NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

REMOVAL OF COHERENT
EXTREMELY LOW FREQUENCY (ELF)
BACKGROUND NOISE BY
ADAPTIVE NOISE CANCELATION

by

Samuel J. Strange

September, 1993

Thesis Advisor: Danny G. Farley
Co-Advisor: Amir-H. Najmi

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE 93 09 23	3. REPORT TYPE AND DATES COVERED Master's Thesis

4. TITLE AND SUBTITLE
REMOVAL OF COHERENT EXTREMELY LOW FREQUENCY (ELF) BACKGROUND NOISE BY ADAPTIVE NOISE CANCELATION

5. FUNDING NUMBERS

6. AUTHOR(S)
Strange, Samuel J.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE
A

13. ABSTRACT *(maximum 200 words)*

The use of the Sequential Regression Algorithm (SER) to coherently remove background noise from an ELF sensor is presented. The SER algorithm is described for a multi-channel application in order to cancel coherent portions of reference sensors from a primary sensor. The algorithm adaptively accounts for differences between two parallel array platforms for the purpose of coherent subtraction. A section on likelihood ratio detector schemes for detecting narrowband signals is also presented. This work is in support of a submerged ELF sensor array project run by the Johns Hopkins University Applied Physics Lab.

14. SUBJECT TERMS Extremely Low Frequency (ELF), Adaptive Noise Cancelation, Sequential Regression Algorithm (SER)			15. NUMBER OF PAGES 73 16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

Approved for public release; distribution is unlimited.

REMOVAL OF COHERENT EXTREMELY LOW FREQUENCY (ELF) BACKGROUND NOISE BY ADAPTIVE NOISE CANCELATION

by

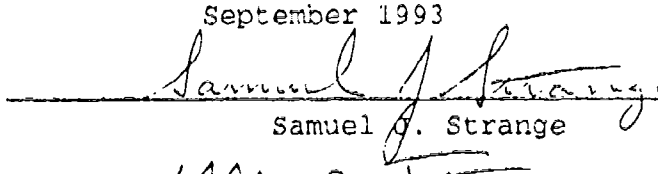
Samuel J. Strange
Lieutenant, United States Navy
B.S., The Citadel, 1986

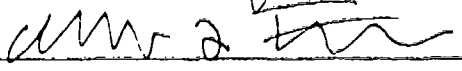
Submitted in partial fulfillment
of the requirements for the degree of

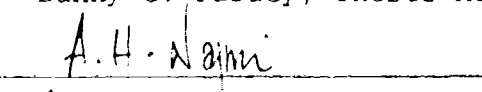
MASTER OF SCIENCE IN SYSTEMS ENGINEERING

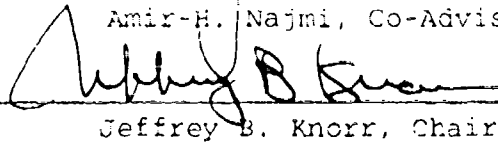
from the

NAVAL POSTGRADUATE SCHOOL
September 1993

Author: 
Samuel J. Strange

Approved by: 
Danny G. Farley, Thesis Advisor


Amir-H. Najmi, Co-Advisor


Jeffrey B. Knorr, Chairman
Electronic Warfare Academic Group

ABSTRACT

The use of the Sequential Regression Algorithm (SER) to coherently remove background noise from an ELF sensor is presented. The SER algorithm is described for a multi-channel application in order to cancel coherent portions of reference sensors from a primary sensor. The algorithm adaptively accounts for differences between two parallel array platforms for the purpose of coherent subtraction. A section on likelihood detector schemes is also presented. This work is in support of a submerged ELF sensor array project run by the Johns Hopkins University Applied Physics Lab.

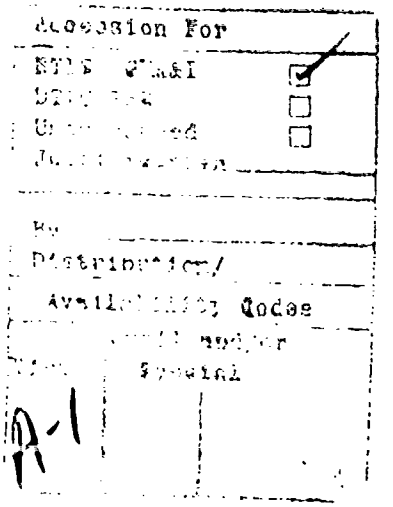


TABLE OF CONTENTS

I. INTRODUCTION

A. BACKGROUND

Extremely low frequency (ELF) geoelectric noise has a highly impulsive nature which makes it very difficult t remove by filtering. The main contributors to the impulsiveness of the ELF spectrum are lightning discharges. The lowest frequencies of the ELF range (1 to 60Hz) have such a low attenuation rate as they propagate, that lighting discharges all over the world must be considered when studying the noise spectrum at a single location. ELF signals have been studied for such applications as power transfer, communications, geophysical surveys, and so forth. The amount of information on the subject is immense and increasing. How to account for the impulsive background electromagnetic fields inevitably present in any system operating in this frequency range remains a serious problem that any system designer must answer.

The background noise of the lower ELF frequency range can be thought of as consisting of both local and global phenomena. The global component is comprised of signals that are relatively coherent over extended distances. A detailed recording of data in this frequency range

> ...is characterized by bursts of a few cycles duration which may be ten or more times the mean amplitude between

bursts. This bursty character is apparently coherent over most, if not all, of the Earth.[Ref. 1]

The local phenomenon is comprised of those components that are not coherent over extended distances and are assumed to be products of the local environment of the detector. A set of parallel detectors separated by some distance can be coherently combined to remove the global portion of the background noise and thus lower the noise floor.

B. OBJECTIVE

The Applied Physics Lab (APL) at Johns Hopkins University has for years studied and made measurements of the background geoelectric noise on widely separated pairs of electrodes. The purpose of this thesis is to describe techniques of adaptive filtering that the author applied to data collected by APL, in order to determine the amount of noise cancellation. The techniques are implemented using software developed during an extended visit to the Applied Physics Lab by the author. In a related effort, software will also be presented implementing appropriate detection methodology for the detection of narrowband signals embedded in the noise.

C. ORGANIZATION

The thesis begins with a brief discussion of the constituents of this ELF noise to demonstrate the difficulty in modeling of the spectrum and to provide supporting background information. The basic concepts of the adaptive

filtering will be presented as a lead in to the adaptive algorithm to be used to cancel the noise. These introductory sections will be followed by a discussion of stability and convergence concerns for the adaptive algorithm in this type of data environment. The detection discussion is then presented as a continuation of the processing required for the detection of narrowband signals. The adaptive algorithm is then used in the detector scheme to demonstrate performance of the developed software. A final section will summarize the work performed, draw relevant conclusions about the work and propose possible future studies that could be derived from this work.

II. THE ELF BACKGROUND

Many studies have been conducted in an attempt to develop an understanding of what actually constitutes the natural and man-made background noise spectrum in the one to sixty hertz range. The sources of the energy in this frequency range can be primarily attributed to atmospheric disturbances and man-made power transfer systems. A major problem with these sources of noise is that they are highly irregular in both magnitude and duration of the noise signal. Attempts to classify and assign specific characteristics to these sources have proven difficult. A brief history of the research conducted in these areas is provided in the following two subsections.

A. ATMOSPHERIC NOISE

The source of the majority of noise in the one to sixty hertz frequency range is lightning discharges. The lower ELF range discussed here also contains what is called the Schumann Range. The Schumann range takes its name from the early theoretical work of the German scientist W. O. Schumann. Schumann theorized that the cavity created by the surface of the Earth and the Ionosphere has naturally occurring resonant frequencies. He took this research further to derive a set of harmonic resonant equations for calculating these frequencies.

The work of Schumann has been carried further with the actual measurement of the Schumann resonance frequencies in field experiments. In 1959 and 1960, H. L. von König substantiated the presence of Schumann resonances by observing the waveforms in the output of a narrow band amplifier. The experimental presence of Schumann resonances created a large quantity of literature on the subject as scientists attempted to understand the phenomenon. Once a foundation of the propagation properties of the Earth-Ionosphere cavity was established, the search for the source of the excitation began.[Ref. 2]

The most prominent excitation source of the Earth-Ionosphere cavity is the cloud to ground lightning stroke. The number of similar strokes present in any one flash observed during a thunderstorm is a random variable that normally ranges from two to twelve. Since the number of strokes is generally too high to measure individual waveforms, a more applicable measure is the power spectral density[Ref. 3]. The frequency spectrum of this noise was resolved in sufficient detail to estimate the quality factor, Q, of the earth-ionosphere cavity by M. Balser and C. A. Wagner in 1960. The Q of a cavity resonator is a measure of the bandwidth of the resonator[Ref. 4]. The source of the noise, as theorized by H. Raemer in 1961, is the response of the earth-ionosphere cavity to electrical discharges created in thunderstorms all over the world. Raemer attempted to

model the ionosphere in an attempt to reproduce mathematically the measured response of Balser and Wagner. Although Raemer's attempt fell short of its goal, it began the process of refinement of a working model of the ionosphere that continues today.[Ref. 2]

The early work performed by these scientists brought out the difficulty in attempting to model this naturally occurring phenomenon due to its randomness. Parameters must be estimated by some method in order to model this random process. The highly variable nature of these parameters makes the task of modeling the noise nearly impossible. Examples of these varying parameters are diurnal variations in the ionosphere, twenty-four hour variations in the ionosphere, seasonal variations in the locations of thunderstorm regions, and the randomness of individual lightning strikes. These are only a few examples of the parameters that must be considered in order to model the atmospheric noise.

B. MAN-MADE NOISE

The presence of atmospheric noise is not the only concern that must be addressed when discussing background noise in the one to sixty hertz range. Many components in today's world emit unshielded interference that falls in this range. These components are the result of either faulty equipment or poor

design. Examples of this noise source are power lines that are used to distribute alternating current through populated areas.

Large ferric objects can create fields around them that can be detected. The strength of these fields can be well below the previously mentioned sources of noise but can be discernable in certain situations. Motion of these large magnetic objects in the Earth's natural magnetic field is the source of a small field.

Motion of the detector can create an increase in the background noise. Any slight motion of the detector can create misleading information and in most cases, an increase in the noise signal. The background ELF spectrum can be thought of as the summation of three orthoganal vectors. Relative motion between these three vectors and the detector creates a false signal that can raise the noise floor. This type of noise can be compensated for by the use of motion sensor.[Ref. 5]

Other sources of man-made noise can be found in the detector equipment itself. Improper shielding of cables and electronic devices used in the design of the detector and its processing components can lead to elevated noise floors. Unlike naturally occurring atmospheric noise, these sources can be traced and eliminated. The removal of this noise can come in two forms, elimination of the source or isolation of

the source from the detector. The fact that man-made noise sources exist must not be overlooked when attempting to process signals at the output of the detector.

III. ADAPTIVE FILTERS

A. GENERAL TERMINOLOGY

The ELF background noise has been described as a highly impulsive, non-stationary random process. Some of the causes of these variations have been briefly described above. The use of traditional filter designs may not be the optimum method for reduction of the noise floor for reception of low strength signals in the one to fifty hertz range. Since their introduction, the use of adaptive filters has shown promise in combating exactly this type of scenario. To understand the application of adaptive filters in this case, a brief discussion of relevant terms must be conducted.

An adaptive filter is by definition a filter that has the capability to adjust, by itself, a set of design parameters that are based on estimated statistical characteristics of the signal to be filtered. This process of adjusting the filter to the present situation alleviates the need to have *a priori* knowledge of the relevant signal characteristics. The previous chapter described how difficult it is to derive a set of statistical characteristics for the ELF noise environment. If *a priori* knowledge of the relevant signal characteristics were known, an optimum filter could be designed. The goal of adaptive filtering is to estimate these statistical parameters

and to refine the estimation through the use of an algorithm. If an appropriate algorithm can be found, then after a sufficient number of iterations, the adaptive filter should converge to the optimum filter. The use of adaptive filters requires the designer to look for the optimum algorithm for estimating the relevant signal characteristics instead of attempting to discover the actual signal characteristics.[Ref. 6]

Adaptive filters are divided into two basic categories, open-loop and closed-loop. An open-loop configuration is usually a two stage process. The first stage is used to "learn" the statistics of the relevant signal. The results of the first stage are then used in the second stage to compute the filter parameters required using a nonrecursive algorithm. In contrast, the closed-loop configuration uses only one stage to develop the filter parameters: the relevant signal characteristics are not explicitly estimated but are inferred from a recursive algorithm that updates the filter parameters directly as each new data point is obtained. The adaptive filter gains additional knowledge of the signal characteristics during each iteration. The gain in knowledge results in an improvement of the filter parameters which are adjusted and their performance is used as an input to the next iteration of the algorithm. The adaptive filter discussed in this thesis falls in the closed loop category. The advantage of closed-loop over open-loop is that the requirement for only

one stage generally corresponds to a less expensive configuration based on its simpler implementation.[Ref 6]

The description of the filters that will be analyzed in this study carries terminology that is pertinent to all filters, not just adaptive filters. Filters are implemented in either continuous-time or discrete-time which defines the form of the relationship between the input and output of the filter. The filters that will be studied are of the discrete-time version which means that the filter may be described by a difference equation. The structure of the filters to be studied are referred to as finite-impulse response (FIR), tapped-delay-line, or transversal filters. This description means that the filter's output relies only on the past and present values of the input. The advantage of a FIR filter is its inherent stability.[Ref 6]

The algorithm that is used to update the filter parameters is generally named after some of its operating characteristics. The algorithm used here is called the Sequential Regression Algorithm or SER[Ref. 7]. This algorithm uses an iterative approach to approximating the weighting factors required for a linear combiner. Figure 1 shows a basic diagram of a linear combiner.

This system uses simultaneous samples of the input signal at time index k, to produce an output signal, y_k, by a linear operation with weighting factors, w. A second way to physically interpret the linear combiner is to look at it in

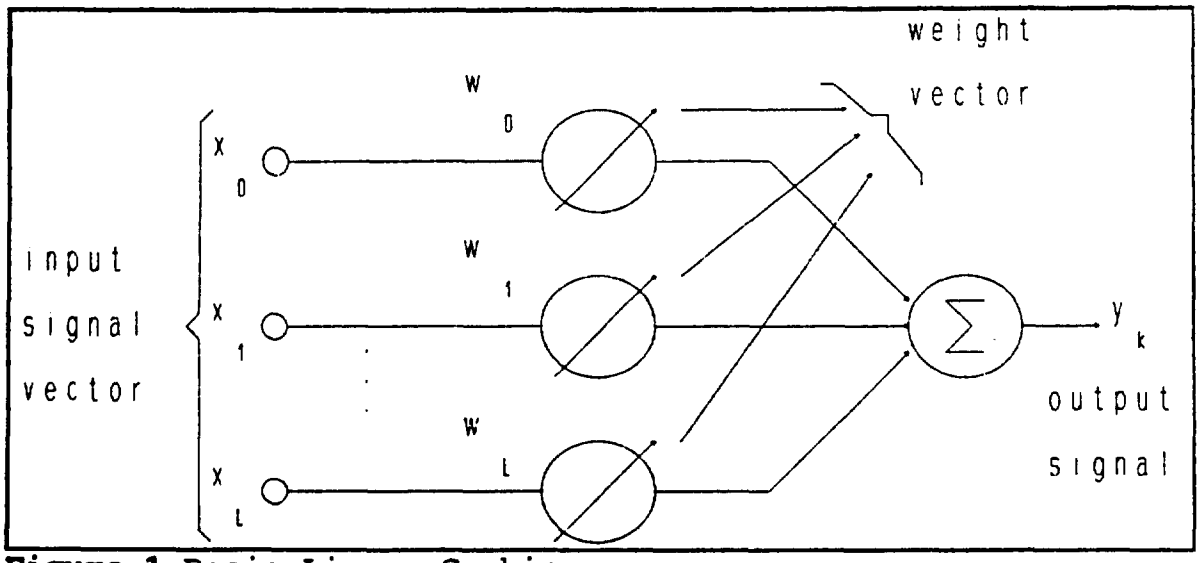


Figure 1 Basic Linear Combiner

the form of a transversal filter. Figure 2 shows how this system is designed: $\mathbf{X}_k$ corresponds to the last L samples of x at time index k. The weighting factors are dual indexed on

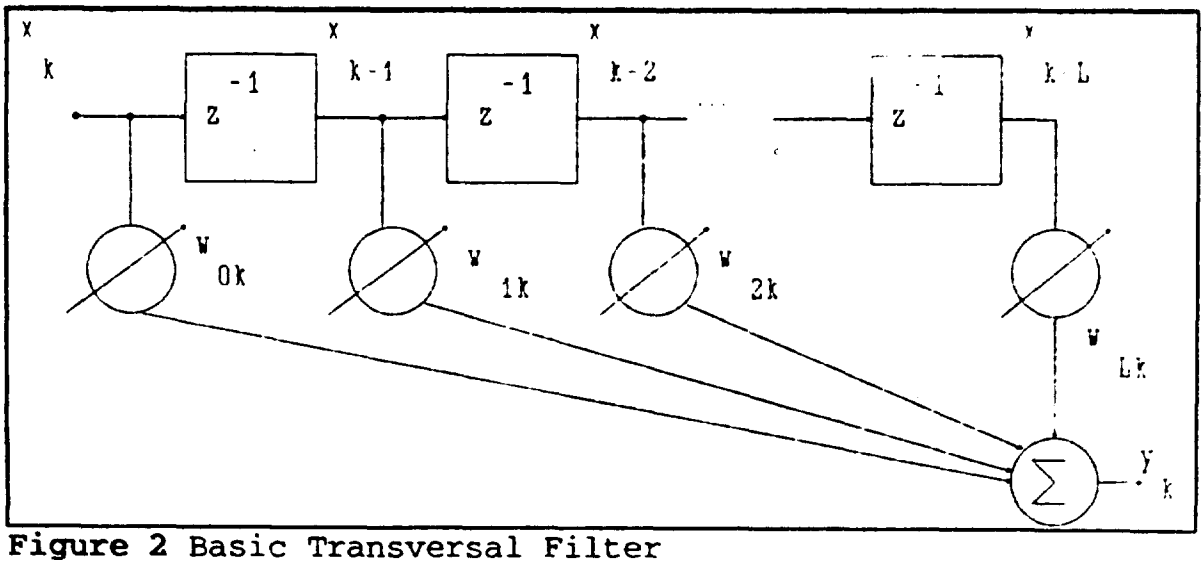


Figure 2 Basic Transversal Filter

the time delay, 1, followed by the time index, k. The weight factor, w_{2k}, corresponds to the weight factor for the sample taken two samples ago from the present sample time, k.

B. THEORY AND DESIGN

1. Configuration

The filter to be developed here is a combination of the previous two figures. The transversal filter is applied to a set of multiple sensor outputs to form an input vector that is N x L data points long, where L is the number of separate inputs or sources, and N is the length of the transversal filter. There are two reasons for using a filter design of this nature. The first is the ability to use more than one source of input, allowing for a better spectral sampling of the background ELF energy. The second is the ability to use the previous N samples thereby smoothing the output since it is not dependent on instantaneous values which may vary greatly from sample to sample. Both of these ideas are beneficial to the overall performance of the filtering process.

The output of the filter, y_k, can be expressed in terms of the input matrix and the weighting factors matrix. To do this, some defining equation must be expressed for the input matrix, $\mathbf{X}_k$,

$$\mathbf{X}_k = [X_{1k} \;\; X_{2k} \;\; \ldots \;\; X_{Lk}]^T \tag{1}$$

and weight factors matrix, $\mathbf{W}_k$,

$$\mathbf{W}_k = [W_{1k} \;\; W_{2k} \;\; \ldots \;\; W_{Lk}]^T \tag{2}$$

where each sensor's input vector is

$$X_{1k} = [x_1(k)\ x_1(k-1)\ x_1(k-2)\ \ldots\ x_1(k-(N-1))] \qquad (3)$$

and weighting factor vector is

$$W_{1k} = [W_{1k}\ \ W_{1(k-1)}\ \ W_{1(k-2)}\ \ \cdots\ \ W_{1(k-(N-1))}]. \qquad (4)$$

The first subscript of Equations 3 and 4, labeled 1, 2, 3, ..., L, indicates the source of the sample. The second subscript, k, indicates the time of the sample. A total of N sample values from each sensor are used for each calculation of y_k. A physical model for the filter to be studied is shown in Figure 3. The scalar output value, y_k in Figure 3, is defined as:

$$y_k = \mathbf{X}_k^T \mathbf{W}_k = \mathbf{W}_k^T \mathbf{X}_k. \qquad (5)$$

These basic equations will become the starting point for the development of an operating filter system for the removal of the ELF background noise. It is preferable that the weighting factors, $\mathbf{W}_k$, cause the output, y_k, to be an estimate of some desired signal, d_k. The concept of the filter is to use the weighting factors to model the differences between signals originating from the same source but sensed at various locations. The inputs, x_i in Figure 3, represent samples of a signal taken at different locations. The desired signal, d_k, is also a sample of the same background source signal. The difference between the desired signal and the output of the filter is expressed as the error in the estimate, ϵ_k.

$$\varepsilon_k = d_k - y_k \tag{6}$$

In the system designed here, the estimate error is the output sought for analysis. If the weight factors are set properly, they will theoretically remove all signals that are coherent between the different sample locations. The goal of

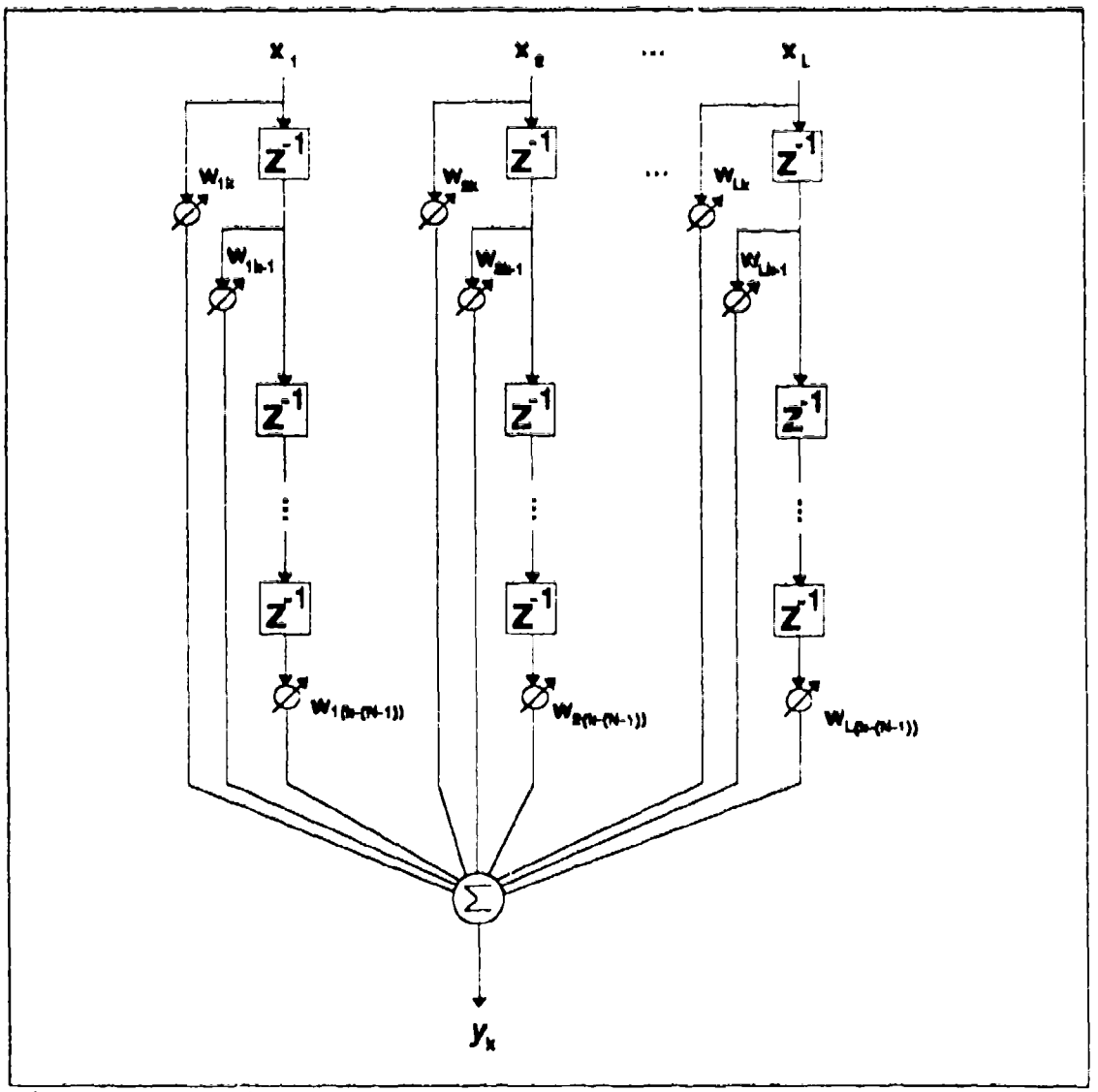


Figure 3 Multi-Channel Transversal Filter

such a system is to remove the correlated background noise from a single sensor thus enhancing signals that are particular to that sensor alone. If the estimated value of y_k can be made to represent the ELF background noise, then the estimate error, ϵ_k, can be assumed to be due only to the signal present in the desired signal. The resulting output from the noise canceler has a much lower noise floor due to the removal of the correlated noise.

2. The Mean Square Error

The filtering process must have a means by which an optimal solution can be defined. The mean-square error will be used to develop a defining equation, called the performance surface, for the optimal solution. Combining Equations 3 and 4, a definition of estimated error in terms of the input vector, the desired signal, and the weighting factors is obtained. The expected value of the squared estimated error, $E(\epsilon_k)^2$, assuming that ϵ_k, d_k, and $\mathbf{X}_k$ are statistically stationary is given as:

$$E[\epsilon_k^2] = E[d_k^2] + \mathbf{W}^T E[\mathbf{X}_k \mathbf{X}_k^T]\mathbf{W} - 2E[d_k \mathbf{X}_k^T]\mathbf{W} \qquad (7)$$

The variables x_k and d_k are assumed to be dependent, and in this case both contain the same ELF background. The equation for the mean square error, MSE, associated with the estimate is thus defined in terms of the auto-correlation function of the input vector and the cross-correlation function between the input vector and the desired signal.

The auto-correlation and the cross-correlation functions will be represented by **R** and **P** respectively. The elements of these two matrices are constant values if the input vector x_k and d_k are stationary random variables. A mathematical expression for the auto-correlation matrix, **R**, is:

$$\mathbf{R} = E\left[\mathbf{X}_k \mathbf{X}_k^T\right] = E\begin{bmatrix} X_{1k}^2 & X_{1k}X_{2k} & \cdots & X_{1k}X_{Lk} \\ X_{2k}X_{1k} & X_{2k}^2 & \cdots & X_{2k}X_{Lk} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ X_{Lk}X_{1k} & X_{Lk}X_{1k} & \cdots & X_{Lk}^2 \end{bmatrix} \tag{8}$$

and the cross-correlation matrix, **P**, is:

$$\mathbf{P} = E[d_k \mathbf{X}_k] = E[d_k X_{1k} \quad d_k X_{2k} \quad \cdots \quad d_k X_{Lk}]^T \tag{9}$$

Using these two definitions in Equation 7 results in the following definition of the mean square error in the k^{th} estimate of y:

$$\text{MSE} \triangleq J_{es} = E[\varepsilon_k^2] = E[d_k^2] + \mathbf{W}^T\mathbf{R}\mathbf{W} - 2\mathbf{P}^T\mathbf{W} \tag{10}$$

The optimum solution for the selection of the weighting factors is the condition where the mean square error is at a minimum.

3. Minimizing The Error

The minimum mean square error condition can be solved for by taking the gradient of the MSE with respect to the weighting factors and setting it equal to zero. The use of

the minimum value of the MSE as the optimal value is based on the assumption that the weighting factors are only designed to model that portion of the input vector that is correlated with the desired signal. If the error is at the minimum value, then the filter is removing the maximum amount of correlated data between the input and the desired signals. The system being evaluated is based on the assumption that only the desired signal contains both noise and a signal of interest while the input vector is composed of a pure noise signal alone. Under this assumption, the only correlated data between the two input data samples is the noise. The desire is to remove as much noise as possible and thus to remove the maximum amount of correlated data between the input vector and the desired signal. It is therefore assumed that the filter is working at its optimum value when the weighting factors meet the criteria of minimizing the mean square error of the output.

4. The Optimum Weighting Vector

The optimum weight factor will be represented by the symbol $\mathbf{W}$. The two conditions that must be met in order for the value of J to be minimized are:

$$\nabla_{\mathbf{W}} J_{ss}|_{\mathbf{W}=\mathbf{W}^{o}} = \mathbf{0} \qquad (11)$$

and

$$\frac{\partial^2 J_{ss}}{\partial w_i \partial w_j} > 0. \tag{12}$$

These two equations state that when the weight vector is at an optimum the performance function is at a local minimum since the slope is zero and the curvature is upward. The gradient of the performance function is evaluated to be:

$$\begin{aligned} \nabla = \nabla_{\mathbf{W}} J &= \nabla_{\mathbf{W}} E[d_k^2] - 2\nabla_{\mathbf{W}}(\mathbf{W}^T\mathbf{P}) + \nabla_{\mathbf{W}}(\mathbf{W}^T\mathbf{R}\mathbf{W}) \\ &= -2\mathbf{P} + 2\mathbf{R}\mathbf{W} \end{aligned} \tag{13}$$

Setting the gradient to zero and solving for the optimum weighting factors results in:

$$\mathbf{W}^o = \mathbf{R}^{-1}\mathbf{P} \tag{14}$$

This equation is sometimes referred to as the Weiner-Hopf equation where the weight factor is referred to as the Weiner weight vector. The input autocorrelation matrix, **R**, can be inverted if it is "positive definite," i.e., if:

$$\mathbf{V}^T\mathbf{R}\mathbf{V} > 0 \tag{15}$$

where **V** is any non-zero vector. When the matrix **R** is "positive definite" then the optimum weight factors can be solved for directly if the complete statistical properties of the auto-correlation and cross-correlation functions are known. The statistical properties are not known in this case so further development is required to obtain an estimate of the optimum weight factors.

The second condition that ensures that the error estimate is at a minimum is that the partial derivative of the gradient with respect to the weighting factors is greater than zero. To prove this the we take the partial derivative of Equation 3 with respect to $\mathbf{W}$:

$$\frac{\partial^2 J}{\partial w_i \partial w_j} = 2\mathbf{R} \tag{16}$$

This result again shows that the input autocorrelation matrix must be "positive definite" in order to solve for the optimum weighting factors.

The performance surface can now be expressed in terms of the minimum error possible, the weighting functions, and the autocorrelation of the input vector. The minimum mean square error possible can be found for the case when the weighting vector is at its optimum value. The optimum weighting factor is used to solve for the minimum error by substituting $\mathbf{W}^{\circ}$ into the minimum error equation.

$$\begin{aligned} J_{min} &= E[d_k^2] + [\mathbf{W}^{\circ}]^T\mathbf{R}\mathbf{W}^{\circ} - 2\mathbf{P}^T\mathbf{W}^{\circ} \\ &= E[d_k^2] + [\mathbf{R}^{-1}\mathbf{P}]^T\mathbf{R}\mathbf{R}^{-1}\mathbf{P} - 2\mathbf{P}^T\mathbf{R}^{-1}\mathbf{P} \\ &= E[d_k^2] - \mathbf{P}^T\mathbf{R}^{-1}\mathbf{P} \\ &= E[d_k^2] - \mathbf{P}^T\mathbf{W}^{\circ} \end{aligned} \tag{17}$$

The minimum error can now be used as the base line by which all other errors are found. The error at any time is the minimum error with the addition of uncertainty in the components used to derive the minimum error. Thus:

$$J = J_{min} + (\mathbf{W}-\mathbf{W}^o)^T\mathbf{R}(\mathbf{W}-\mathbf{W}^o) \qquad (18)$$

The expression can be shown to equal that of Equation 10 by the following proof. Using the fact that $(\mathbf{A}-\mathbf{B})^T$ in general can be expressed as $\mathbf{A}^T-\mathbf{B}^T$:

$$J = J_{min} + [\mathbf{W}^o]^T\mathbf{R}\mathbf{W}^o + \mathbf{W}^T\mathbf{R}\mathbf{W} - \mathbf{W}^T\mathbf{R}\mathbf{W}^o - [\mathbf{W}^o]^T\mathbf{R}\mathbf{W} \qquad (19)$$

since all terms are scalars, the transpose is equal to the scalar value, thus the last two terms are equal. Substituting for the minimum error and the optimum weighting vector:

$$\begin{aligned} J &= E[d_k^2] - \mathbf{P}^T\mathbf{R}^{-1}\mathbf{P} + [\mathbf{W}^o]^T\mathbf{R}\mathbf{W}^o + \mathbf{W}^T\mathbf{R}\mathbf{W} - 2\mathbf{W}^T\mathbf{R}\mathbf{W}^o \\ &= E[d_k^2] - \mathbf{P}^T\mathbf{R}^{-1}\mathbf{P} + \mathbf{P}^T\mathbf{R}^{-1}\mathbf{R}\mathbf{R}^{-1}\mathbf{P} + \mathbf{W}^T\mathbf{R}\mathbf{W} - 2\mathbf{W}^T\mathbf{R}\mathbf{R}^{-1}\mathbf{P} \\ &= E[d_k^2] + \mathbf{W}^T\mathbf{R}\mathbf{W} - 2\mathbf{W}^T\mathbf{P} \\ &= E[d_k^2] + \mathbf{W}^T\mathbf{R}\mathbf{W} - 2\mathbf{P}^T\mathbf{W} \end{aligned} \qquad (20)$$

This result validates Equation 18 by showing that it equivalent to Equation 10. An important note about this expression of the performance surface is that it is quadratic with respect to the weighting vector.

5. Iterative Calculation of W_k

The optimum weight factors can be arrived at iteratively by use of a gradient search of the performance function. What this means is that after each estimate of the optimum weight vector, $\mathbf{W}_k$, the slope or gradient of the performance function, in the direction of the optimum weighting vector, is used to derive the next weighting vector, $\mathbf{W}_{k+1}$. This method ensures that each estimation of the optimum

weight factors is at least as close or closer to the optimum value than the previous estimate. Multiplying Equation 13 on the left by ½ $\mathbf{R}^{-1}$ and combining the result with Equation 14, we obtain:

$$\mathbf{W}^{o} = \mathbf{W} - \frac{1}{2}\mathbf{R}^{-1}\nabla \tag{21}$$

To convert this into an iterative process, the optimum solution is considered to be the value of the weighting vector for the next iteration.

$$\mathbf{W}_{k+1} = \mathbf{W}_{k} - \frac{1}{2}\mathbf{R}^{-1}\nabla_{k} \tag{22}$$

This equation is referred to as the Newton Method algorithm for determining the weighting vector.

The algorithm takes only one iteration to arrive theoretically at the optimum solution for the values of the weighting vector. This algorithm would be the ideal method to use if exact solutions to the gradient term and the inverse correlation of the input vector were known exactly. This is not generally the case and an estimate for the gradient must be used. To compensate for the lack of knowledge of the gradient at iteration k, a constant u, will be used instead of the factor ½ in Equation 22. The requirements on the range of values that u can take on are:

$$0 < \mu < 1 \qquad (23)$$

The effect of μ on the solution at each iteration is that it affects the rate of convergence of the algorithm. A discussion of the attributes and further limitations on the value of μ will be covered later. A general form of the Newton's method of gradient search is therefore:

$$\mathbf{W}_{k+1} = \mathbf{W}_k - \mu \mathbf{R}^{-1} \nabla_k \qquad (24)$$

This expression is ideal under the following three conditions:

1. $\mu = \frac{1}{2}$
2. Exact knowledge of the gradient vector, ∇_k.
3. Exact knowledge of the (unchanging) matrix $\mathbf{R}^{-1}$.

These conditions, unfortunately, are not normally attainable. The value of μ is selected between 0 and $\frac{1}{2}$ so that the algorithm is overdamped and can accommodate fluctuations in the matrix $\mathbf{R}$. The compensation for this selection of μ is that the algorithm takes longer to obtain its optimum solution. A second modification to the value of μ is to base its value on the average eigenvalue, λ_{av}, so that μ is replaced by $\mu\lambda_{av}$.

The second condition pertaining to the exact solution to the gradient of the error surface is, in general, never attainable. The gradient must therefore be estimated in some manner. The estimate of the gradient that will be used is

based on the square of the present value of the error signal and not the true ensemble average as it has been up to now.

$$\hat{\nabla}_k = \frac{\partial e_k^2}{\partial \mathbf{W}_k} = 2e\frac{\partial e_k}{\partial \mathbf{W}_k} \tag{25}$$

From the definition of the error signal:

$$\frac{\partial e_k}{\partial \mathbf{W}_k} = \frac{\partial [d_k - \mathbf{X}_k^T \mathbf{W}_k]}{\partial \mathbf{W}_k} = -\mathbf{X}_k \tag{26}$$

which leads to the following estimate of the gradient:

$$\hat{\nabla}_k = -2e_k\mathbf{X}_k \tag{27}$$

With these two relaxations from the three ideal conditions, the algorithm is now in a form known as the Newton/LMS algorithm[Ref 7]:

$$\mathbf{W}_{k+1} = \mathbf{W}_k + 2\mu\lambda_{av}\mathbf{R}^{-1}\mathbf{X}_k \tag{28}$$

This is an optimum iterative algorithm that is only complete if full knowledge of the input correlation function is known. This is generally not the case and thus a final modification must be done to produce a working estimate to this algorithm. The final modification is an iterative estimate of the inverse of the input correlation matrix, $\mathbf{R}^{-1}$, and is called the Sequential Regression (SER) algorithm.

IV. THE SEQUENTIAL REGRESSION (SER) ALGORITHM

A. DEVELOPMENT

The Sequential Regression (SER) algorithm uses the LMS/Newton algorithm as its basis in conjunction with a method for iteratively estimating the value of the inverse input auto-correlation matrix to produce the weighting vector for the filter coefficients. The development of the algorithm comes from *Adaptive Signal Processing* by Widrow and Stearns[Ref 7]. The core of the algorithm is the use of an estimate for the inverse of the input auto-correlation matrix which reduces computational load with minimal estimation error.

The input auto-correlation matrix is

$$\mathbf{R}=E[\mathbf{X}_k\mathbf{X}_k^T] \tag{29}$$

where the subscript k goes over the entire length of the random process $\mathbf{X}_k$. $\mathbf{R}$ will be estimated iteratively using only a finite or truncated sample of the random process $\mathbf{X}_k$. The estimate of the input auto-correlation matrix is given by:

$$\hat{\mathbf{R}}_k = \frac{1}{k+1}\sum_{l=0}^{k}\mathbf{X}_l\mathbf{X}_l^T. \tag{30}$$

This estimate is unbiased when the input variable, $\mathbf{X}_k$, is a stationary random process. When $\mathbf{X}_k$ is not stationary, it can be shown that the estimate is not very accurate. The estimate

of $\mathbf{R}_k$ at each value of k is equally influenced by the previous, k-1, estimates of the input auto-correlation matrix. In order that the most recent estimates of $\mathbf{R}_k$ have more influence over the new estimate, a "fading" memory term, α, is introduced which reduces the significance of the past estimates. The fading memory expression of the input auto-correlation matrix is denoted by $\mathbf{Q}_k$ where:

$$\mathbf{Q}_k = \sum_{l=0}^{k} \alpha^{k-l} \mathbf{x}_k \mathbf{x}_k^T. \tag{31}$$

The value of α is chosen, as a rule of thumb, such that the half life of the exponential function is equal to the length of stationarity of $\mathbf{X}$ [Ref 7]. This rule of thumb leads to the following statements about the value of α.

$$\alpha \approx 2^{(-1/\text{length of stationarity of } \mathbf{x})} \tag{32}$$

$$0 < \alpha < 1 \tag{33}$$

The value of the fading memory factor over k iterations is:

$$\sum_{l=0}^{k} \alpha^{k-l} = \frac{1-\alpha^{k+1}}{1-\alpha} \tag{34}$$

and thus the estimate of the input auto-correlation matrix becomes:

$$\hat{\mathbf{R}}_k = \frac{1-\alpha}{1-\alpha^{k+1}} \mathbf{Q}_k = \frac{1-\alpha}{1-\alpha^{k+1}} \sum_{l=0}^{k} \alpha^{k-l} \mathbf{x}_l \mathbf{x}_l^T. \tag{35}$$

This estimate is exact for the condition where $\mathbf{X}_k$ is constant. If the input variable, $\mathbf{X}_k$, is stationary, α is one, and if the limit of Equation 35 is taken as α approaches one, the result is in agreement with Equation 30.

The derivation of the SER algorithm begins with the estimate of the input auto-correlation matrix. Using this estimate in the equation for the optimum weighting factors results in

$$\hat{\mathbf{R}}_k \mathbf{W}_k = \hat{\mathbf{P}}_k. \tag{36}$$

The definition for the estimate of the cross-correlation vector, $\hat{\mathbf{P}}_k$, is, similarly to Equation 35,

$$\hat{\mathbf{P}}_k = \frac{1-\alpha}{1-\alpha^{k+1}} \sum_{l=0}^{k} \alpha^{k-l} d_l \mathbf{X}_l. \tag{37}$$

Using Equations 35 and 37 in 36 gives:

$$\mathbf{Q}_k \mathbf{W}_k = \sum_{l=0}^{k} \alpha^{k-l} d_l \mathbf{X}_l. \tag{38}$$

The SER algorithm attempts to estimate the next value of the weighting vector, $\mathbf{W}_{k+1}$, based on the present values of $\hat{\mathbf{R}}_k$ and $\hat{\mathbf{P}}_k$. To do this, Equation 38 becomes:

$$\begin{aligned} \mathbf{Q}_k\mathbf{W}_{k+1} &= \sum_{l=0}^{k-1} \alpha^{(k-1)-l} d_l \mathbf{X}_l + d_k \mathbf{X}_k \\ &= \alpha \mathbf{Q}_{k-1}\mathbf{W}_k + d_k \mathbf{X}_k \\ &= (\mathbf{Q}_k - \mathbf{X}_k\mathbf{X}_k^T)\mathbf{W}_k + d_k \mathbf{X}_k. \end{aligned} \tag{39}$$

The last line of Equation 39 is the result of the following relationship:

$$\mathbf{Q}_k = \alpha \mathbf{Q}_{k-1} + \mathbf{X}_k \mathbf{X}_k^T. \tag{40}$$

The value of d_k can be replaced by recalling Equation 5 and 6 which state that

$$d_k = e_k + y_k = e_k + \mathbf{X}_k^T \mathbf{W}_k \tag{41}$$

thus Equation 39 becomes

$$\begin{aligned} \mathbf{Q}_k\mathbf{W}_{k+1} &= (\mathbf{Q}_k - \mathbf{X}_k\mathbf{X}_k^T)\mathbf{W}_k + (e_k + \mathbf{X}_k^T\mathbf{W}_k)\mathbf{X}_k \\ &= \mathbf{Q}_k\mathbf{W}_k + e_k\mathbf{X}_k. \end{aligned} \tag{42}$$

Multiplying on the left by $\mathbf{Q}_k^{-1}$, gives an iterative method for calculating the weighting factor vector

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mathbf{Q}_k^{-1} e_k \mathbf{X}_k. \tag{43}$$

The expression for $\mathbf{W}_{k+1}$ in Equation 43 is similar to the LMS/Newton method derived earlier. With this in mind, an approximation for the LMS/Newton algorithm can be made using the above derivation. The SER algorithm used as an approximation of the LMS/Newton method is given as:

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \frac{2\mu\lambda_{ave}(1-\alpha^{k+1})}{1-\alpha}\mathbf{Q}_k^{-1}e_k\mathbf{X}_k. \tag{44}$$

The algorithm requires a method of iteratively solving for the value of $\mathbf{Q}_k^{-1}$ in order to be complete.

The method to iteratively solve for $\mathbf{Q}_k^{-1}$ begins by pre-multiplying Equation 40 by $\mathbf{Q}_k^{-1}$, post-multiplying by $\mathbf{Q}_{k-1}^{-1}$, and then by $\mathbf{X}_k$, to obtain:

$$\mathbf{Q}_k^{-1}\mathbf{Q}_k\mathbf{Q}_{k-1}^{-1}\mathbf{X}_k = \mathbf{Q}_k^{-1}\alpha\mathbf{Q}_{k-1}\mathbf{Q}_{k-1}^{-1}\mathbf{X}_k + \mathbf{Q}_k^{-1}\mathbf{X}_k\mathbf{X}_k^T\mathbf{Q}_{k-1}^{-1}\mathbf{X}_k \tag{45}$$

which reduces to,

$$\begin{aligned}\mathbf{Q}_{k-1}^{-1}\mathbf{X}_k &= \alpha\mathbf{Q}_k^{-1}\mathbf{X}_k + \mathbf{Q}_k^{-1}\mathbf{X}_k\mathbf{X}_k^T\mathbf{Q}_{k-1}^{-1}\mathbf{X}_k \\ &= \mathbf{Q}_k^{-1}\mathbf{X}_k(\alpha + \mathbf{X}_k^T\mathbf{Q}_{k-1}^{-1}\mathbf{X}_k)\end{aligned}. \tag{46}$$

The quantity in parentheses is a scalar and can be divided out while the quantity $\mathbf{X}_k^T\mathbf{Q}_{k-1}^{-1}$ is also multiplied on the right to give

$$\frac{\mathbf{Q}_{k-1}^{-1}\mathbf{X}_k\mathbf{X}_k^T\mathbf{Q}_{k-1}^{-1}}{\alpha + \mathbf{X}_k^T\mathbf{Q}_{k-1}^{-1}\mathbf{X}_k} = \mathbf{Q}_k^{-1}\mathbf{X}_k\mathbf{X}_k^T\mathbf{Q}_{k-1}^{-1}. \tag{47}$$

The first lines of Equation 46 and 47 can be combined and re-arranged to form,

$$\mathbf{Q}_k^{-1} = \frac{1}{\alpha}\left[\mathbf{Q}_{k-1}^{-1} - \frac{(\mathbf{Q}_{k-1}^{-1}\mathbf{X}_k)(\mathbf{Q}_{k-1}^{-1}\mathbf{X}_k)^T}{\alpha + \mathbf{X}_k^T(\mathbf{Q}_{k-1}^{-1}\mathbf{X}_k)}\right]. \tag{48}$$

An iterative method of calculating $\mathbf{Q}_k^{-1}$ is now available and is only based on its previous value, $\mathbf{Q}_{k-1}^{-1}$, and the present value of the input, $\mathbf{X}_k$. The value in parentheses is used three

times, it will be calculated separately and represented by $\mathbf{S}_k$.

The SER algorithm is now completely derived as an iterative method of approximating the LMS/Newton method of adaptive filtering. The following summary shows the process by which each new value of the weighting factor vector is calculated.

$$
\begin{aligned}
\alpha &\approx 2^{(-1/\text{length of stationarity of } \mathbf{X})} \\
\varepsilon_k &= d_k - \mathbf{W}_k \mathbf{X}_k^T \\
S &= \mathbf{Q}_{k-1}^{-1} \mathbf{X}_k \\
\gamma &= \alpha + \mathbf{X}_k^T \mathbf{S} \\
\mathbf{Q}_k^{-1} &= \frac{1}{\alpha}\left(\mathbf{Q}_{k-1}^{-1} - \frac{1}{\gamma}\mathbf{S}\mathbf{S}^T\right) \\
\mathbf{W}_{k+1} &= \mathbf{W}_k + \frac{2\mu\lambda_{ave}(1-\alpha^{k+1})}{1-\alpha}\mathbf{Q}_k^{-1}\varepsilon_k\mathbf{X}_k \\
0 < \mu &< \frac{1}{\lambda_{max}} \; ; \; \mu\lambda_{ave} < 1
\end{aligned}
\tag{49}
$$

The values that must be carried forward from one iteration to the next are $\mathbf{Q}_{k-1}^{-1}$ and $\mathbf{W}_k$. These represent the previous value for the estimate of the inverse input auto-correlation matrix and the present estimate for the weighting factor vector.

The output of the algorithm is the error, ε_k, between the present value of the desired signal and estimate of the desired signal:

$$\text{OUTPUT} = \varepsilon_k = d_k - \mathbf{W}_k \mathbf{X}_k^T \quad (50)$$

This value is saved after each iteration in a vector representing the uncorrelated portion of the desired signal and noise vector of the primary channel with the noise only vectors of the reference channels at the time index k, $\mathbf{X}_k$. [Ref 7]

B. INITIALIZATION

The SER algorithm assumes that the value of the inverse input auto-correlation matrix and the weight factors vector is known from the previous step. At some point in the past, an estimate for these terms must have been made so that a starting point can be established. The required accuracy of the estimates depends on the length of time that the algorithm will be operating and the earliest time that accurate data is required.

The driving force behind the accuracy of the initial estimate is the speed with which the algorithm can converge to an optimal solution. The speed of convergence is dependent on two elements of the algorithm. The first element is the initial value of the inverse input auto-correlation matrix, $\mathbf{Q}_0^{-1}$. The significance of this initial element is that the further the initial estimate is from the true value, the further the algorithm has to adapt in order to approximate the optimum solution. The second element is the step size, $2\mu\lambda_{min}$,

which specifies how far the algorithm can progress towards the optimum value in one iteration. The step size will be discussed in the next section but for now it shall be assumed to be a constant in this discussion of the initial conditions.

The use of a large constant multiplied by the identity matrix as the initial value of the Q_k^{-1} has been sufficiently argued by Lee [Ref. 8].

$$Q_0^{-1} = q_0 I \tag{51}$$

As discussed by Lee, the use of a large constant for q is the proper selection for the case where the random process is stationary. Widrow and Stearns argue that this same choice for the initial value of q_0, in a non-stationary condition, is also appropriate[Ref 7]. The point is made that if a priori knowledge of the random process allows for a better estimate of the value of Q_0^{-1}, then that information should be used. The method of initialization for this study is to use Equation 51 with q_0 equivalent to 100.

The initial values used in the weighting factor vector are assumed to be zero. This assumption is made since knowledge of the process before time zero is unknown. The use of a 32 tap delay filter also allows for an estimate of the initial full weight factor vector but this is done so at the expense of the first 32 output data points. This is done by expanding the number of elements in the weight factor vector as each of the first 32 points are read into the algorithm. The

algorithm is applied only to the data that is available; therefore, the first time that each element of $\mathbf{W}_k$ is non-zero is at time index 31. This corresponds to the time when the first 32 data points are input into the algorithm. The result of this initialization is that the first real data point to be considered an output of the SER method used here is at $k = 31$.

V. STABILITY AND CONVERGENCE

A. THE SINGLE WEIGHT CASE

The SER algorithm contains values that can be set to affect the rate at which the algorithm achieves the optimum solution. A simple case where only one weight factor is concerned will be used to demonstrate the relationships involved in the selection of the convergence rate factors. The single weight iterative process using a gradient search method like the SER algorithm can be represented by:

$$w_{k+1} = w_k + \mu(-\nabla_k) \tag{52}$$

The expression for the single weight gradient at time index k is given by:

$$\begin{aligned} \nabla_k &= \left.\frac{d\mathcal{J}}{dw}\right|_{w=w_k} \\ &= \left.\frac{d[\lambda(w-w^o)^2]}{dw}\right|_{w=w_k} \\ &= 2\lambda(w_k - w^o) \end{aligned} \tag{53}$$

The value of the input auto-correlation matrix, r_{xx} for the single weight case, is replaced by its equivalent eigenvalue, $\lambda = E[x_k^2]$. Combining Equations 52 and 53, an iterative equation for the weight factors is found:

$$w_{k+1} = w_k - 2\mu\lambda(w_k - w^o) \tag{54}$$

for the single weight case. The terms in Equation 54 can be rearranged to combine like terms.

$$w_k = w^o + (1-2\mu\lambda)^k(w_0 - w^o) \tag{55}$$

This expression contains the geometric ratio of successive terms which is also used to define the stability requirements.

Stability is guaranteed if the absolute value of the geometric ratio is less than one, i.e.,

$$|1-2\mu\lambda| < 1 \tag{56}$$

The limits placed on the value of the μ can be expressed in terms of the eigenvalue of the input auto-correlation matrix λ by rearranging Equation 56

$$\frac{1}{\lambda} > \mu > 0 \tag{57}$$

The optimum solution can be seen as the solution of Equation 55 if the limit of w_k as k goes to infinity. This is because the geometric ratio approaches zero as k approaches infinity. The rate at which the algorithm approaches the optimal value is dependent on the value of the geometric ratio.

Figure 4 shows the effect of varying values of the geometric ratio on the rate of convergence. If the value of the geometric ratio can be maintained between zero and one, then the algorithm will always approach the optimal solution from one side and can be considered as overdamped. For a

geometric ratio of zero, the algorithm reaches the optimum value in one iteration from one side and can be considered as critically damped. If the magnitude of the geometric ratio falls between zero and negative one then the algorithm is underdamped.

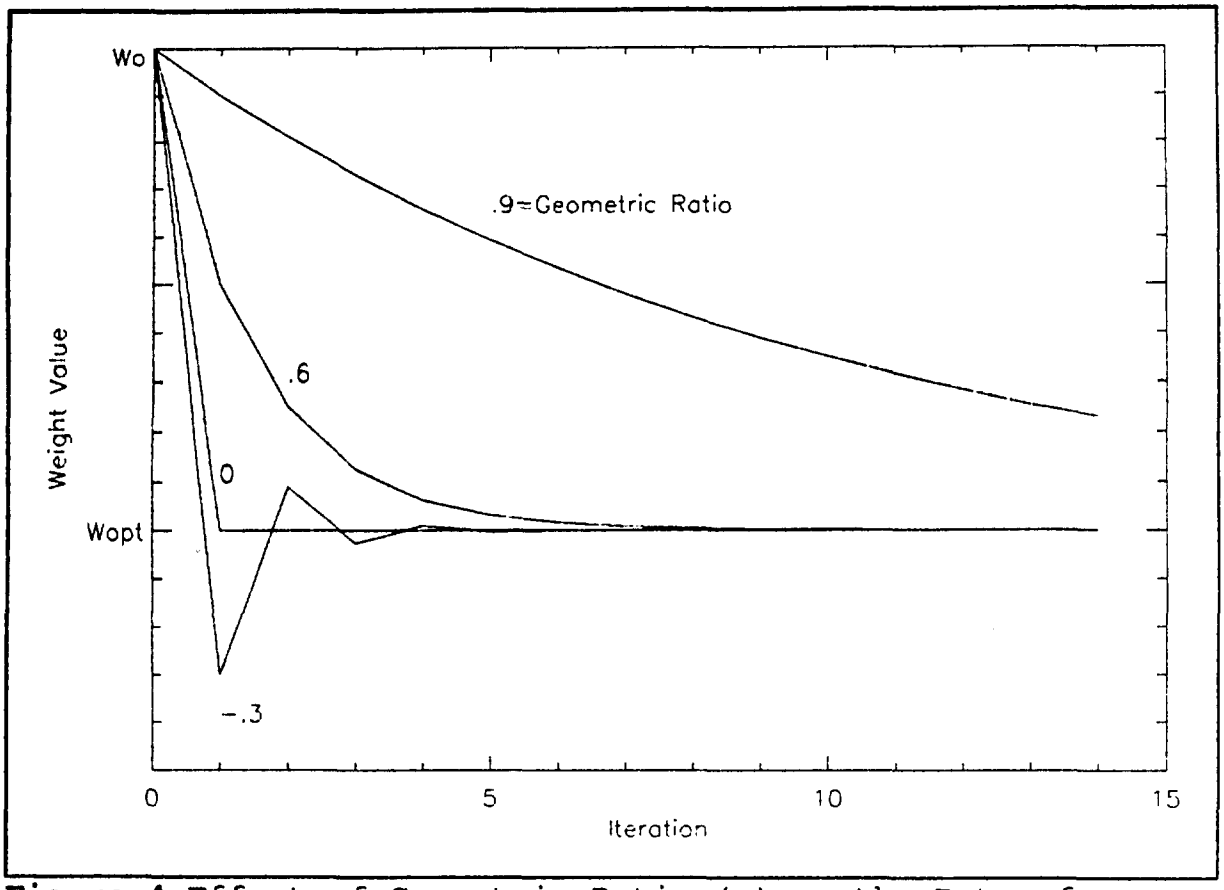


Figure 4 Effect of Geometric Ratio (g) on the Rate of Convergence: overdamped 0<g<1; critically damped g=0; underdamped -1<g<0

B. MULTI-WEIGHT CONDITION

The single-weight condition discussed in the previous section can be expanded to develop an understanding of the multiple weight factor case. The same conditions that applied to the single weight case are true in the multiple weight

case. The difference is that each weighting factor has an associated eigenvalue and thus the condition of stability is now dependent on a set of simultaneous equations that must meet the stability condition.

$$\mathbf{W}_k=\mathbf{W}^o+\begin{bmatrix} (1-2\mu\lambda_0) \\ (1-2\mu\lambda_1) \\ (1-2\mu\lambda_2) \\ \vdots \\ (1-2\mu\lambda_{L-1}) \end{bmatrix}(\mathbf{W}_0-\mathbf{W}^o) \qquad (58)$$

The subscript on λ indicates the index value corresponding to a specific weight value. There are a total of L weighting factors and thus there are L individual eigenvalues, λ.

The value of μ in each of the simultaneous equations is held constant. Since μ is the same for all equations, the eigenvalue that has the largest value will impose the most restrictive conditions on the selection of a value for μ. The new range of values that μ can take on is:

$$\frac{1}{\lambda_{max}}>\mu>0 \qquad (59)$$

The condition stated in Equation 59 gives the multi-weight stability requirements that must be met to ensure that the algorithm will converge to the optimal solution.

The rate at which the algorithm will converge was shown in Figure 4 to be dependent on the value of the geometric ratio. Now that the value of μ is a function of the maximum eigenvalue of the input auto-correlation matrix, **R**, it can be

shown how this affects the rate of convergence. By holding the value of μ constant for each weighting factor, an undesirably slow convergence rate can occur for widely varying values of λ. As shown previously, the algorithm converges in one iteration when μ is equal to $1/2\lambda$ for the single weight case. Using that same criteria, the value of the multi-weight constant μ is $1/2\lambda_{max}$. The use of this criterion ensures that the algorithm falls into the overdamped category since μ is always less than or equal to $1/2\lambda_l$ for $l=[0, 1, 2, \ldots, L]$. The weight factor, w_l, with the smallest value of λ will have the slowest convergence rate since its geometric ratio will be farthest from zero. This is true because the value of $2\mu\lambda$ will be at its minimum and will cause the geometric ratio to approach one. When this condition holds, the algorithm may take excessively long periods of time to converge to an optimal solution.

C. NON-STATIONARITY EFFECTS

The discussion up to this point in dealing with stability and convergence assumed that the process under study was stationary. Under stationary conditions, the values of the eigenvalues of **R** are assumed to be constant. A second key assumption to the development is that the values of the eigenvalues are known. The lack of knowledge of the input auto-correlation matrix, **R**, requires that some estimation be made for the value of μ. The value of μ must be adjusted to

ensure convergence of the algorithm in an environment where the eigenvalues of the auto-correlation matrix are varying at some unknown rate from iteration to iteration.

A method for compensating for the variations in the range of values that μ can assume is to use an averaged eigenvalue, λ_{ave}, for the selection of the constant μ for each iteration. The value of μ is now a constant for a single iteration but is changed from iteration to iteration based on the average eigenvalue. Stability of the algorithm is guaranteed if the value of the product, $\mu\lambda_{ave}$, is between zero and one. It is important to note that the expression $\mu\lambda_{ave}$ is now used only to replace the constant μ used previously. The idea is to scale the average eigenvalue for that iteration by a constant, μ, and use it as the factor for controlling convergence of the algorithm.

The geometric ratio which controls the convergence of the algorithm has now been altered to include a term to compensate for the non-stationarity of the input random process, $\mathbf{X}$. The new expression for the geometric ratio is

$$|1 - 2\mu\lambda_{ave}\mathbf{R}^{-1}| < 1 \qquad (60)$$

The inverse input auto-correlation matrix, $\mathbf{R}^{-1}$, is included in the development of the SER algorithm. The convergence controlling term is thus $\mu\lambda_{ave}$, and must be less than the maximum eigenvalue of $\mathbf{R}$ to ensure stability and convergence. A second condition that can be imposed on the product $\mu\lambda_{ave}$ is

that its value must always be between zero and one. Using the fact that the spread between maximum and minimum eigenvalues can be rather large, the product of $\mu\lambda_{ave}$ must be much less than one.

The ELF background noise processes have shown a tendency to have a wide range between eigenvalues and to slowly change over time. Because of these two conditions, the value of the $\mu\lambda_{ave}$ must be adjusted over time to account for these variations. The chosen method is to estimate the average eigenvalue by taking the average of the square of the input vector, $\mathbf{X}_k$. This value is multiplied by the 1/100 of the inverse of the maximum of the squared input vector.

$$\mu\lambda_{ave} = \frac{1}{100\,[\max(x_1)]^2} \times \frac{\sum_{I=0}^{L} x_I^2}{L} \qquad (61)$$

The use of this estimate for convergence control in the algorithm ensures that the requirements for stability are met for each iteration. The 1/100 term ensures that the maximum value of μ is not exceeded. It was derived empirically with the use of an actual set of ELF background measurements and may require adjustment in a different environment.

The choice of using the present input to the algorithm to get an estimate of the average eigenvalue is used for computational ease with minimal added error. The use of the inverse input auto-correlation matrix should provide a more accurate measure of the eigenvalues; That, however, requires

the inversion of the matrix. The prevention of matrix inversion is the exact reason the SER algorithm is developed, so to obtain the eigenvalues in this way defeats the reason for the algorithm's design, and is therefore counter-productive. The method has proven adequate and reasonably accurate in empirical work up to this point, but does leave room for improvement and further study.

VI. IMPLEMENTATION

A. A PHYSICAL DESCRIPTION

The Applied Physics Lab has two identical sensor platforms submerged in shallow water. Each platform has electrode pairs for detecting the geoelectric ELF background noise. The system has been in operation for over two years collecting samples of the background noise environment.

Ideally, the two platforms are parallel when the platforms are placed on the bottom of the bay. Figure 5 shows the ideal alignment of the platforms. The two platforms are labeled as the east, E, and west, W, platforms. Each platform has a set of orthogonal electrode pairs, labeled X and Y. These electrode pairs are orientated to receive the two orthogonal source vectors, $\mathbf{E}_x$ and $\mathbf{E}_y$, in the horizontal plane. The strongest geoelectric field strengths are in the horizontal plane due to the shift in polarization as a wave crosses the water-air boundary from vertical to horizontal.

If the two platforms are perfectly aligned and the noise recorded on each are identical, the noise in a specific direction, X or Y, can be removed by subtraction. To cancel the x-direction noise simply subtract XW from XE. Because the platforms are not perfectly parallel and the recorded noises are not identical though highly correlated, this method is not

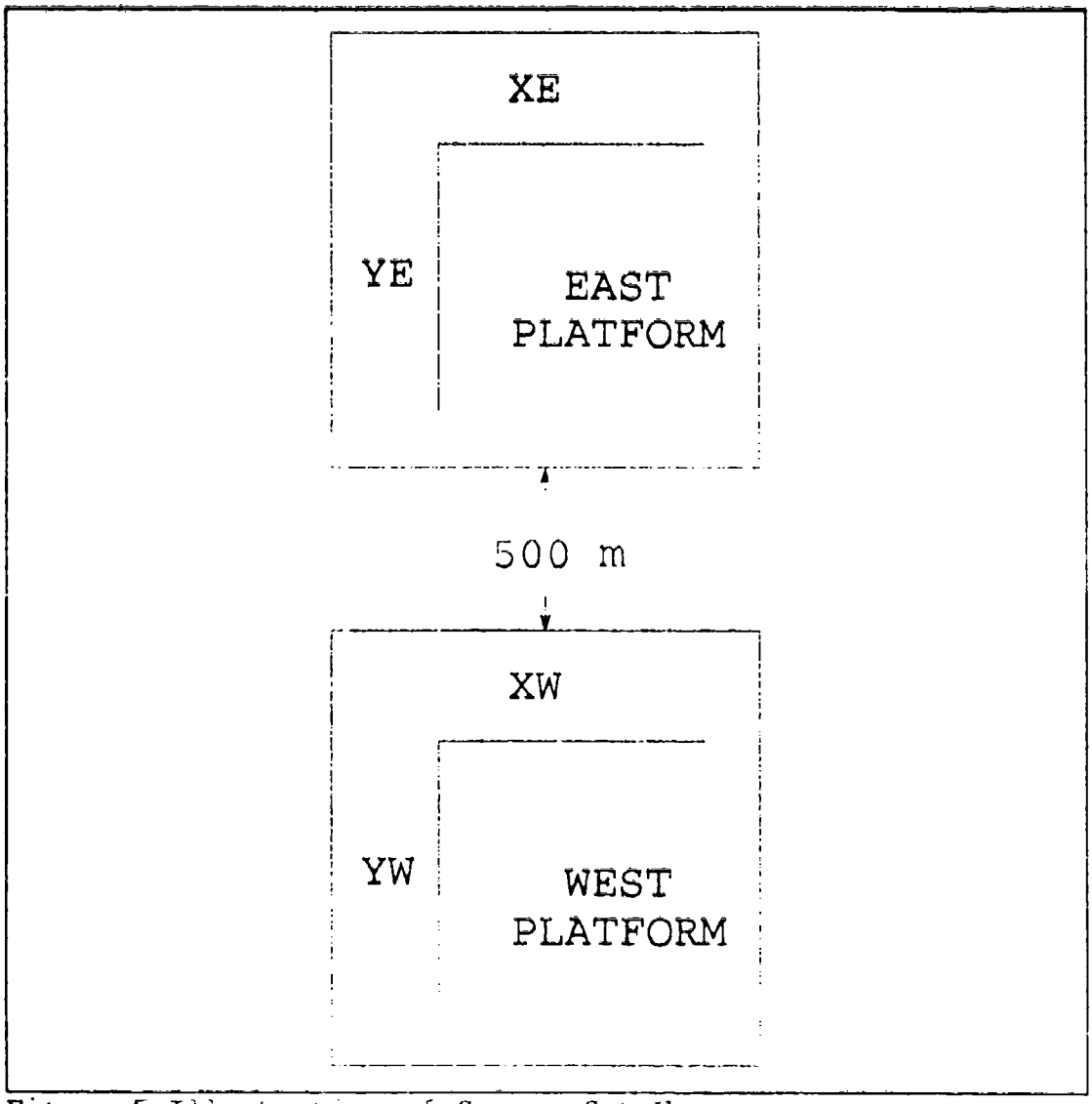


Figure 5 Illustration of Sensor Set-Up

ideal. The actual platforms can be misaligned up to 10 degrees from perfectly parallel. This is due to physical constraints such as bottom contours and mounting technique. The result of such a misalignment is that both the XW and YW sensors of the west platform have components that are coherent with XE of the east platform. If the electrode pairs in

Figure 5 are thought of as vectors, then it is easily seen how misalignment can cause coherency between XE, XW, and YW.

A set of real data is traced through the signal processing system designed to accompany this sensor system. A calibrated source signal is embedded in the signal XE. The path that the received signal follows can be seen in Figure 6. The output

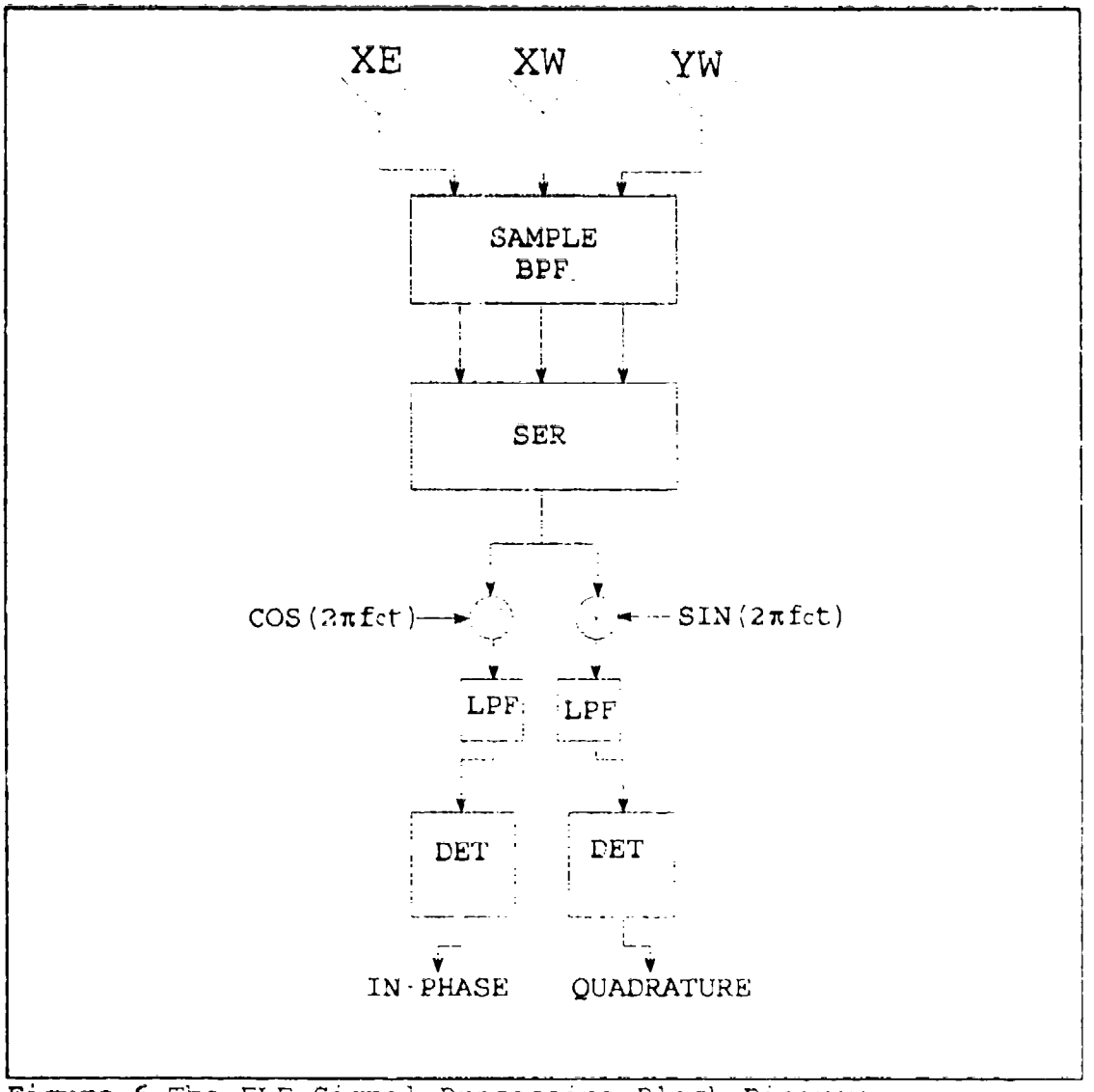


Figure 6 The ELF Signal Processing Block Diagram

of each sensor is sampled at 128 Hz. The data is bandpass filtered and resampled down to 32 Hz for processing. The data is then passed through the SER algorithm where the coherent portions of the west platform are removed from XE. The output of the SER portion of the processor is demodulated and passed through a likelihood ratio detector where a decision can be made as to when the embedded signal is present.

B. INPUTS TO THE SER ALGORITHM

The SER algorithm is used to reduce the ELF background noise detected in the primary submerged electric field sensor by cancellation with the output of the reference parallel electric field sensor. These parallel reference sensors are also submerged and are, ideally, assumed to be located within the same ELF background noise as the primary channel. The primary channel *y* represents sensor XE and the reference channels, *x1* and *x2*, represent XW and YW, respectively. The samples are the actual output of the sensor system used by APL. The samples are taken at the same time without any known signals present. The sampling rate of the data input for the SER algorithm is 32 Hz and the total length of data is 113760 points. These data samples correspond to almost an hour of background noise measurements.

The filter is a 32 point transversal filter that is applied to each channel simultaneously. The length of the transversal filter means that the previous 31 samples of the

sensor are used with the present sample value for each iteration of the algorithm. The goal of the algorithm is to take the output of the two reference channels, *x1* and *x2*, and to generate an estimate of the noise in the primary channel *y*. Figure 7 shows what the primary channel's data looks like in real time. The length of stationarity, used to determine the forgetting factor, α, is assumed to be about 300 data points

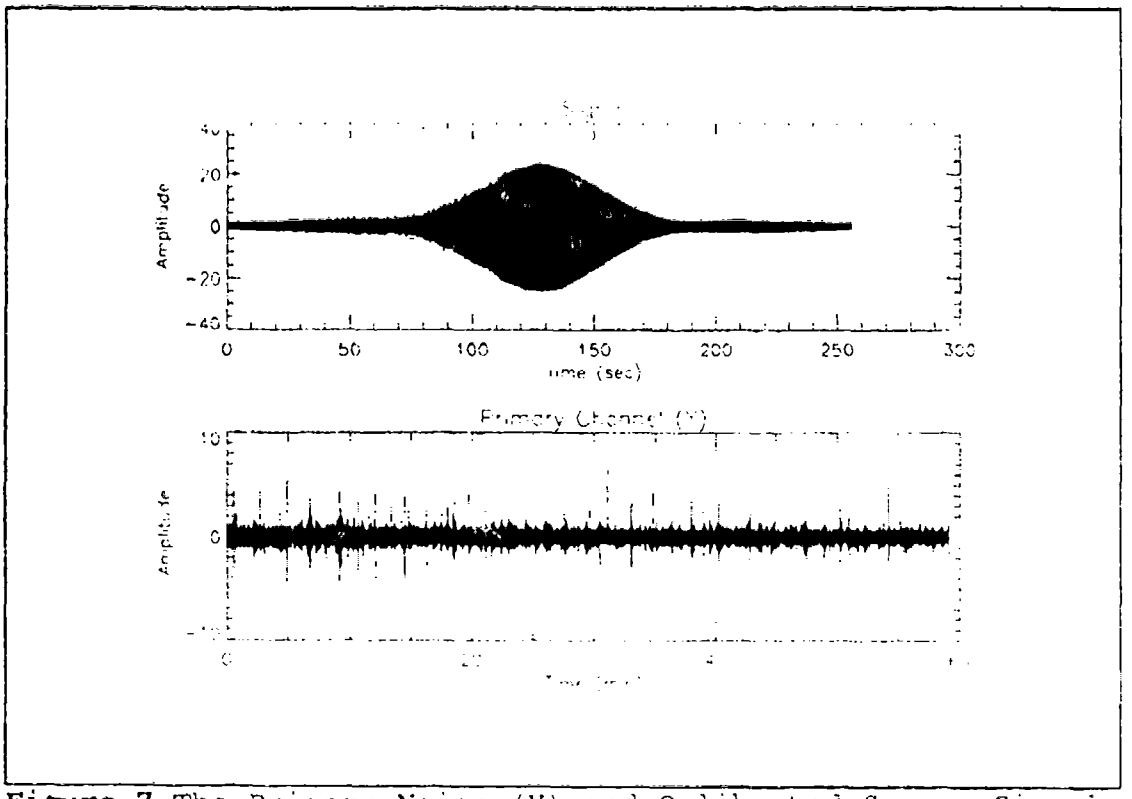


Figure 7 The Primary Noise (Y) and Calibrated Source Signal

long. This value is used based on empirical analysis of the real data.

The effectiveness of the algorithm is analyzed by running a set of scenarios with a known signal inserting in the data

samples. The scenarios consist of manually inserting a 32 Hz signal on a 5 Hz carrier that is 8192 data points long. Figure 7 shows the components of the primary channel signal in a real time representation. The primary channel is processed by the SER algorithm while varying the strength and location of signals embedded in the background noise y.

C. SIGNAL STRENGTH RESULTS

The algorithm is first run with the location of the signal within the noise held constant. The strength of the signal is varied by dividing each data point in the signal vector by a constant. A series of four runs are conducted, each succeeding run has half the signal strength of the preceding run. The initial run uses the signal shown in Figure 7 with the amplitude divided by 500. The signal is indistinguishable prior to processing as seen in Figure 8.

The signals for each run are centered at time 19.2, 32 and 44.8 minutes into the noise sample y. There is no significance to the positions chosen except that the signal data streams were not allowed to overlap. The objective of this first set of runs is to demonstrate how the SER algorithm enhances the ability to distinguish small signals buried in the noise. The data is first processed in the SER algorithm to remove the coherency between the reference channels and the primary channel. The output is then demodulated leaving only the in-phase and quadrature components of the non-coherent

portion of the primary channel. An identical reference set of data is demodulated without passing it through the SER algorithm so that a visual and quantitative analysis of the effects of the SER algorithm can be made. The output of the demodulator is shown in Figures 9 and 10. Note that the envelope of the signal is visible to nearly the strength of 1/2000th, on the third run made.

The output of the demodulator is passed as input into a matched filter detector based on the demodulated signal. The output of this detector is shown in Figures 11 and 12 with the SER and non-SER output for a given signal strength shown

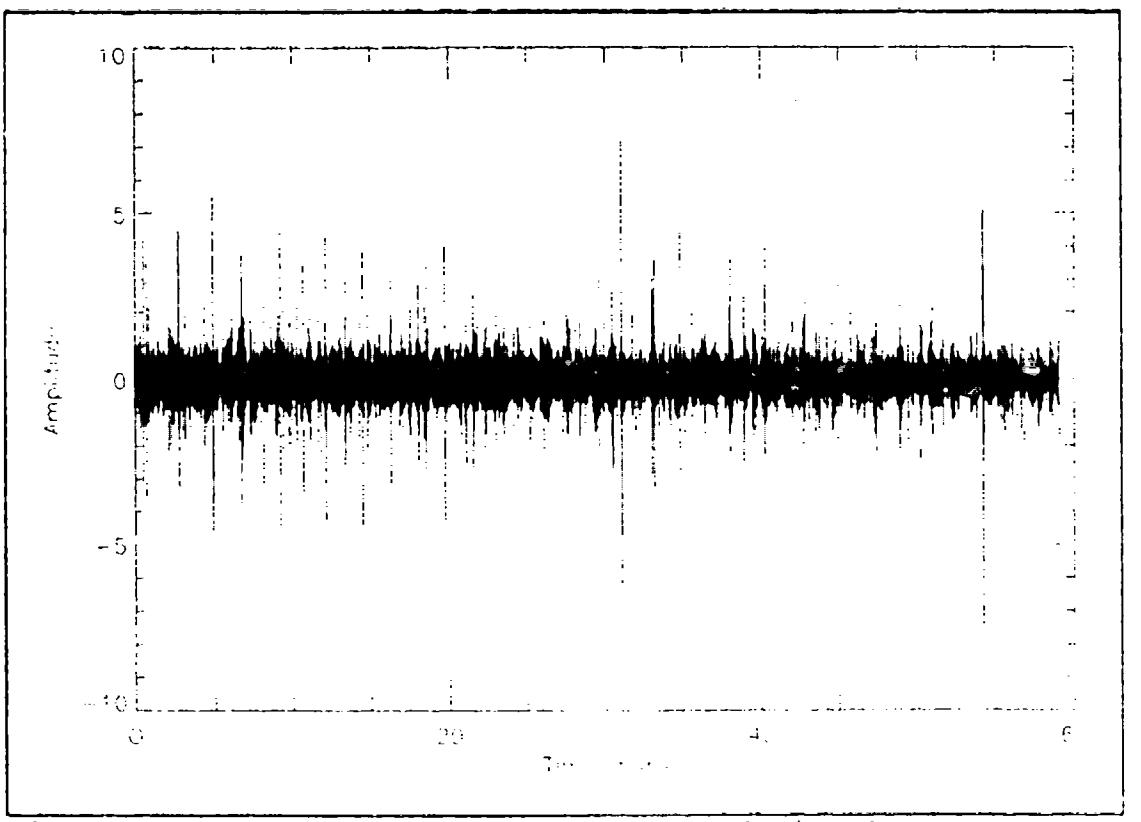


Figure 8 Input Primary Channel Noise And Signal

together. Numerically, the peak value out of the detector is three to five times greater when the SER algorithm is used as compared to when it is not used.

A final numerical analysis is performed by using a primary channel without a signal installed to get a measure of the noise present in the output. The standard deviation of this output of the detector when this noise only vector is used as input is used as an approximation to the noise power. Dividing the outputs of the detector in Figures 11 and 12 by this estimate of the noise power, an estimate of the signal-to-noise-ratio (SNR) can be made. The values of the average peak SNR for the three locations in each run are plotted versus signal strength in Figure 13. The SNR of the SER applied data runs about 6 dB higher than the non-SER applied data.

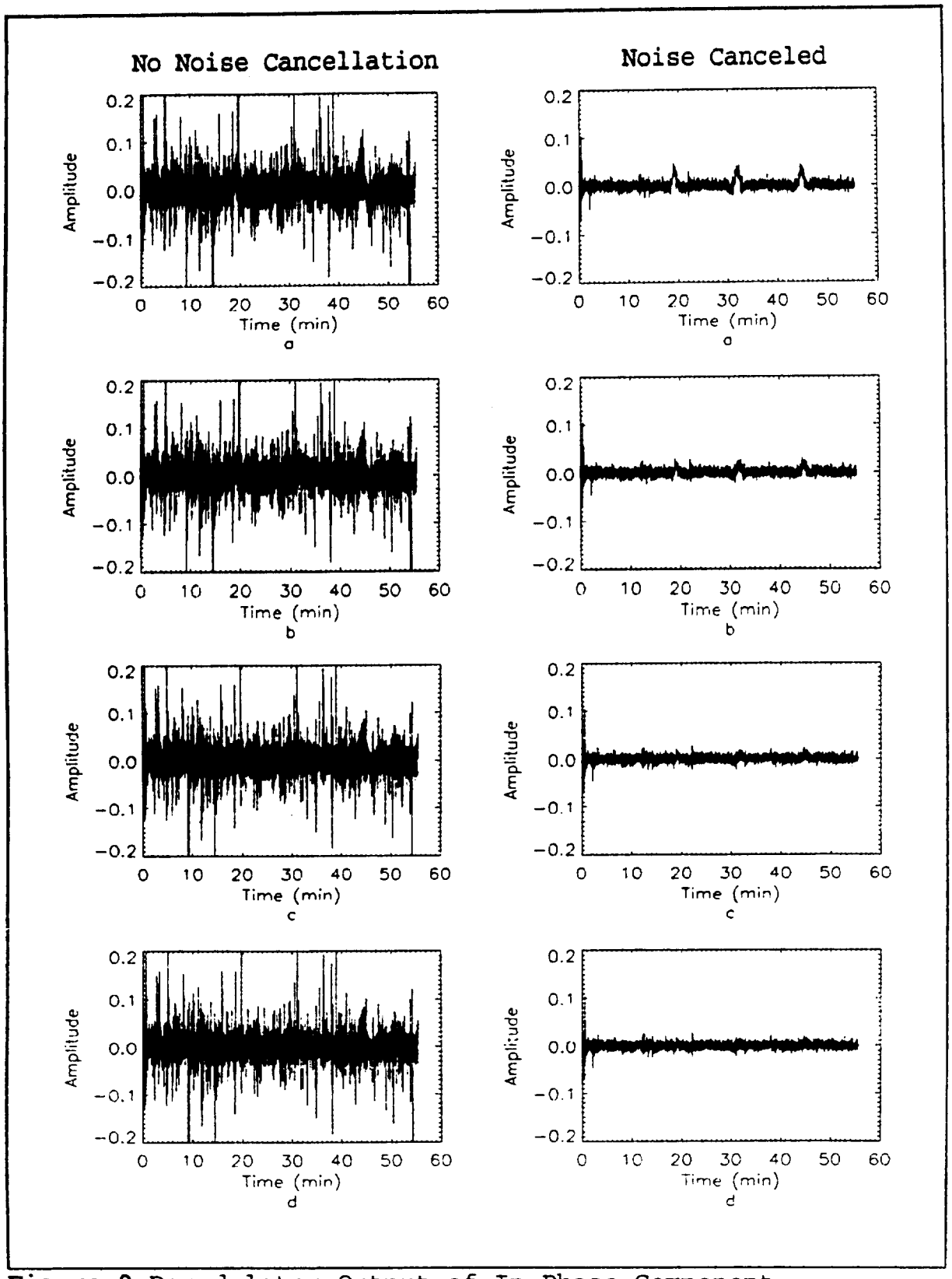


Figure 9 Demodulator Output of In-Phase Component; (a) Signal/500 (b) Signal/1000 (c) Signal/2000 (d) Signal/4000

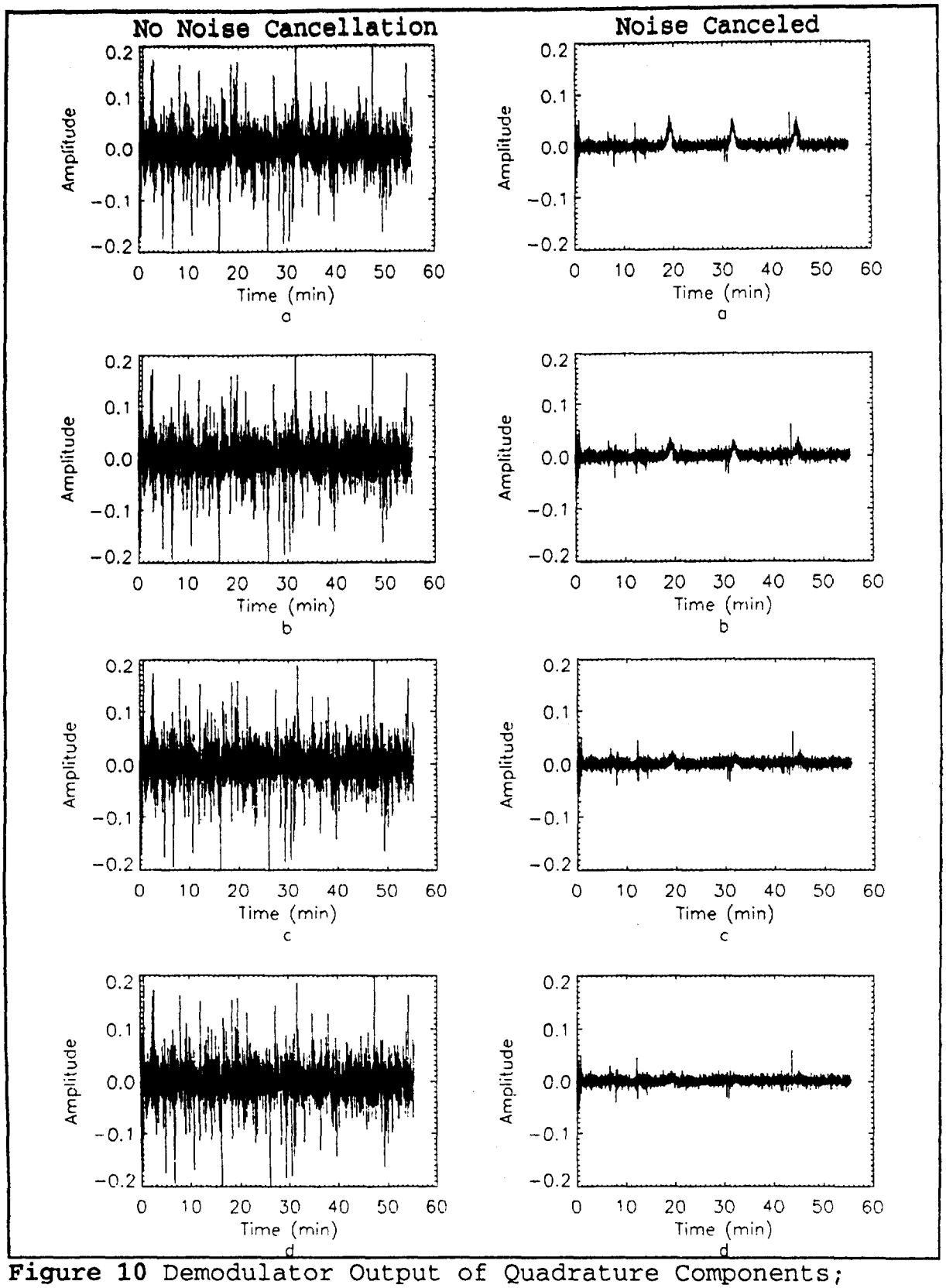


Figure 10 Demodulator Output of Quadrature Components; (a) SIGNAL/500 (b) SIGNAL/1000 (c) SIGNAL/2000 (d) SIGNAL/4000

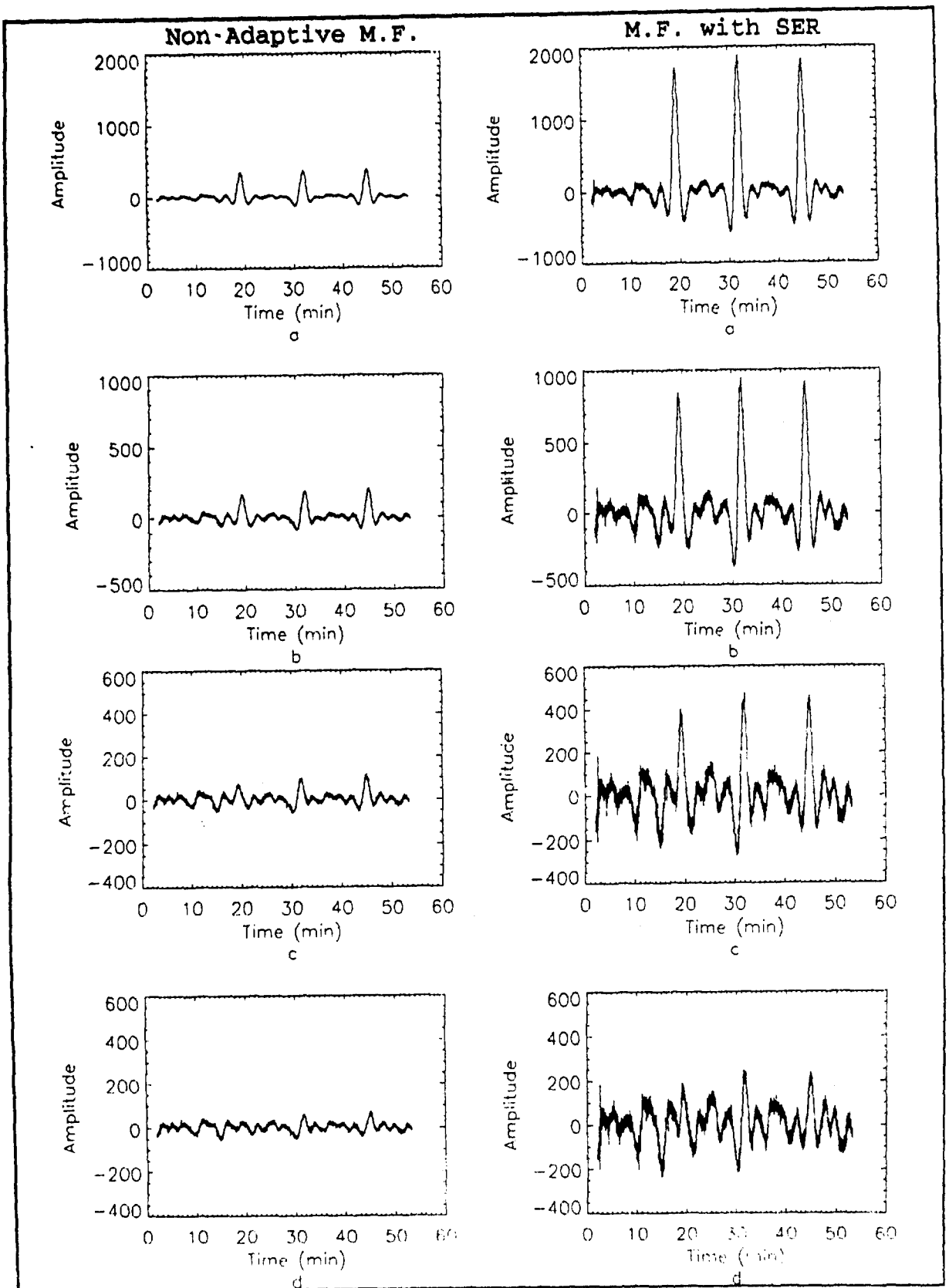


Figure 11 In-Phase Detector Outputs: (a) Signal/500 (b) Signal/1000 (c) Signal/2000 (d) Signal/4000

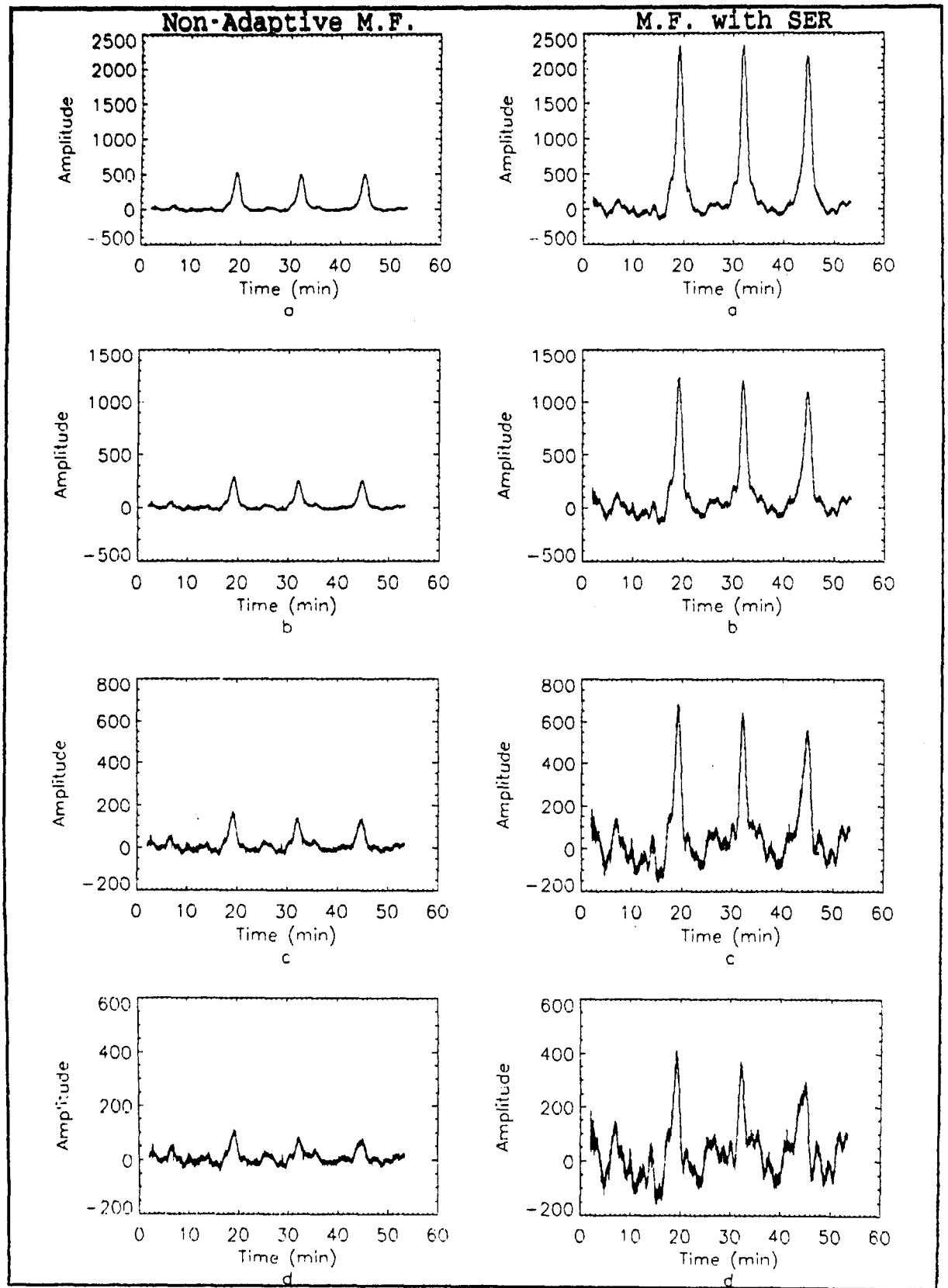


Figure 12 Quadrature Detector Outputs: (a) Signal/500 (b) Signal/1000 (c) Signal/2000 (d) Signal/4000

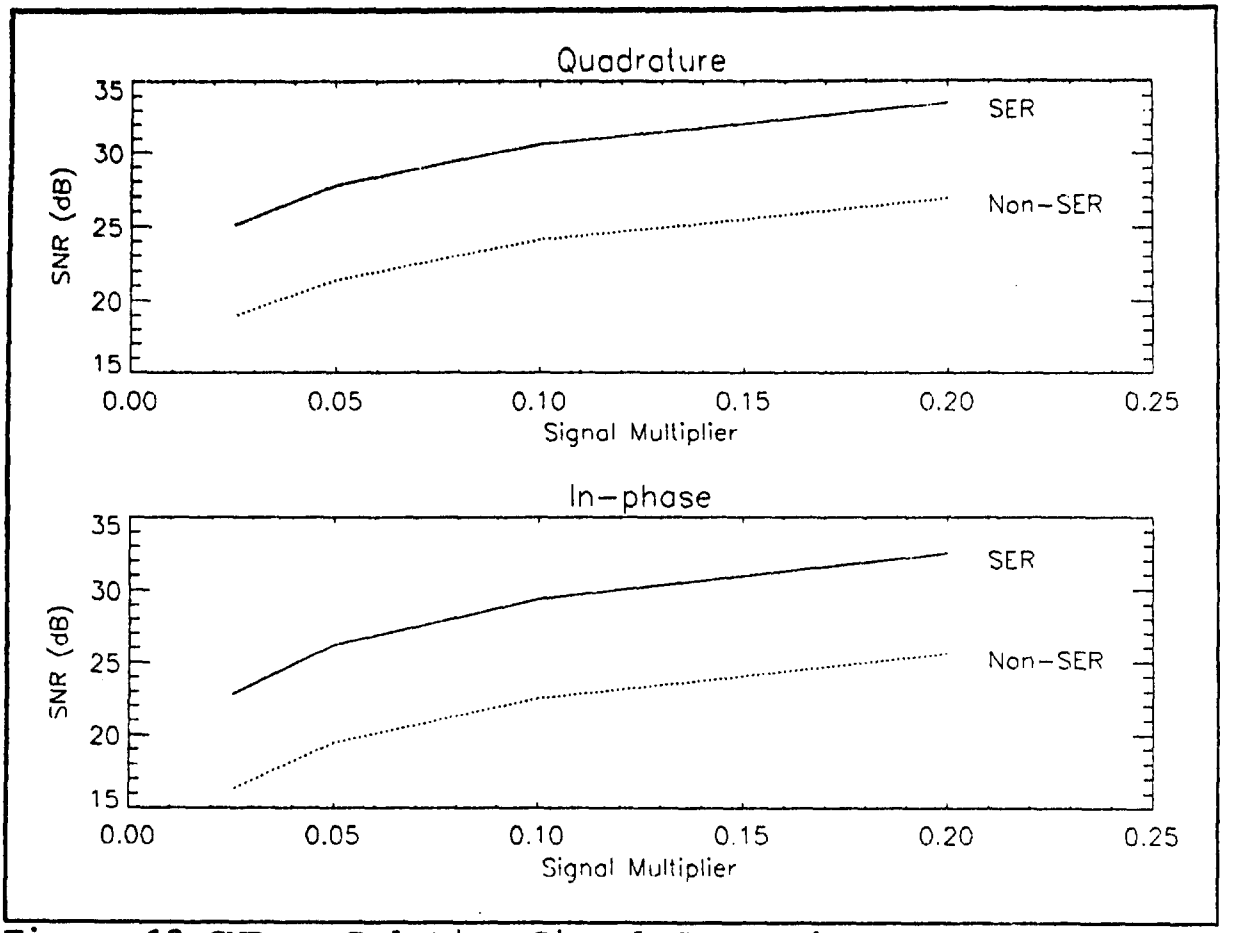


Figure 13 SNR vs Relative Signal Strength

D. DETECTOR THEORY

The process of adaptive noise cancellation is performed for the purpose of increasing the signal-to-noise ratio. A section on the narrowband detection schemes that have been considered in conjunction with the noise canceler are presented since they are integral to the analysis here. Two likelihood ratio detectors are considered for the detection of narrowband signals.

The first detector is designed with the assumption that not only is the signal known, but its peak within a given interval is constant relative to the start of that interval. This last assumption is to assure that the demodulated in-phase and quadrature signals that are used in the likelihood ratio detector match those in the noise canceled and demodulated output of the sensor system described previously. In the second detector, the latter assumption was dropped and the design followed that of an optimal envelope detector. In both cases it was assumed that the residual noise field was gaussian. This assumption, while not strictly true, seems to yield quite reasonable results.[Ref. 9]

The signal is generated by a calibrated source with known qualities. The modelling equation for the narrowband signal is given by:

$$s(t) = A(t)\cos[2\pi f_c t + \theta(t)] \qquad (62)$$

For the first detector, the phase term, θ, is assumed constant within any non-overlapping interval 0 to T. The interval of 0 to T is equivalent to the length of the signal. The purpose of this assumption is so that a Matched Filter can be created for the in-phase and quadrature components by using the demodulated in-phase and quadrature component of the signal. Figure 14 shows the demodulated signals used for the detector.

The first detector is a multi-channel Matched Filter shown in Figure 15. The Matched Filters, h_1 and h_2, are determined

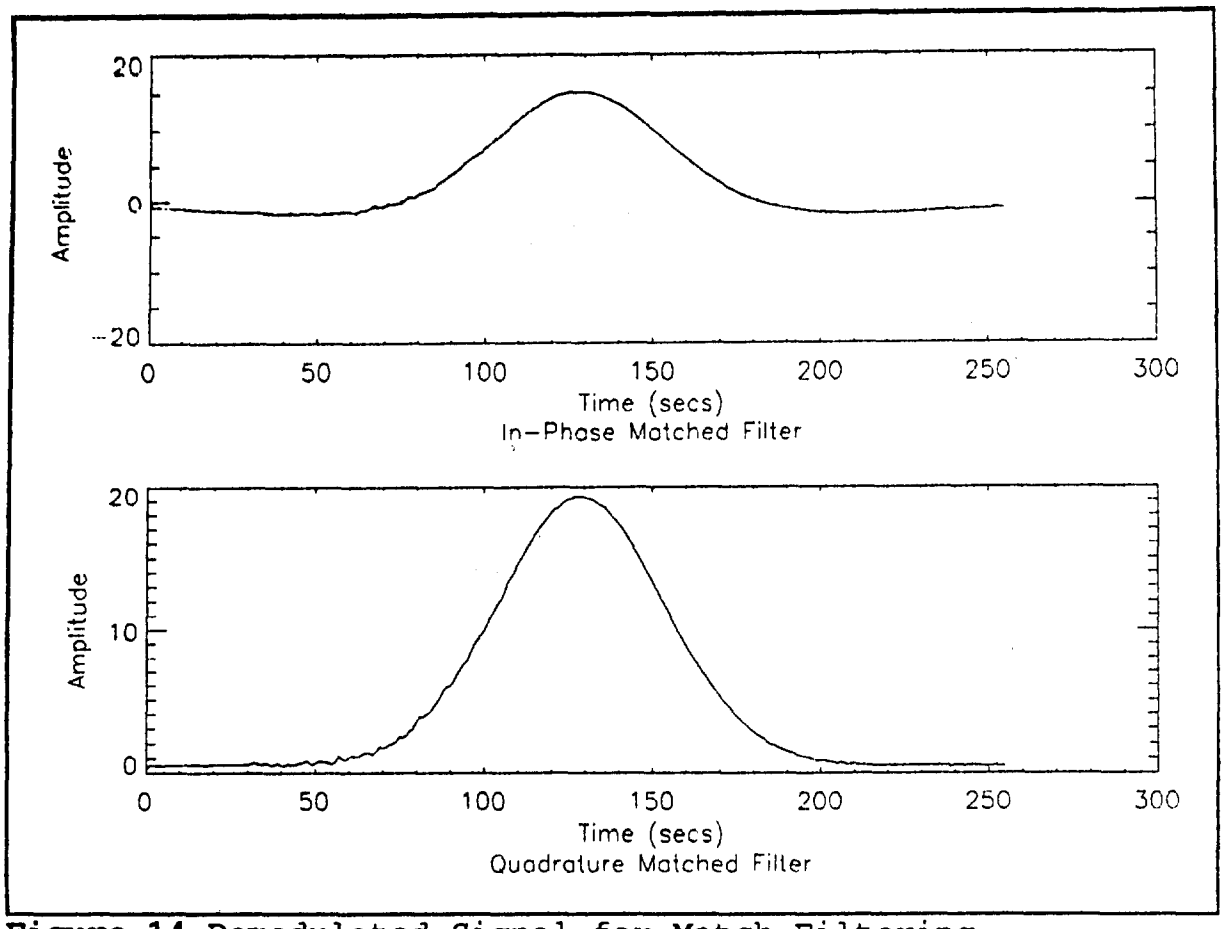


Figure 14 Demodulated Signal for Match Filtering

based on the requirement that the output signal-to-noise ratio is maximized at time T. The corresponding equations are

$$\int_0^T \begin{bmatrix} R_{ii}(t-\tau) & R_{iq}(t-\tau) \\ R_{qi}(t-\tau) & R_{qq}(t-\tau) \end{bmatrix} \begin{bmatrix} h_i(\tau) \\ h_q(\tau) \end{bmatrix} d\tau = \begin{bmatrix} s_i(T-\tau) \\ s_q(T-\tau) \end{bmatrix} \qquad (63)$$

where the quantities R denote the various auto and cross correlations. This detector is the one used for all processing performed in this thesis. The input to the detector is the noise cancelled output of the SER portion of the processor after demodulation.[Ref. 10]

The second detector is an optimal nonlinear envelope detector. The description of its principles and design are

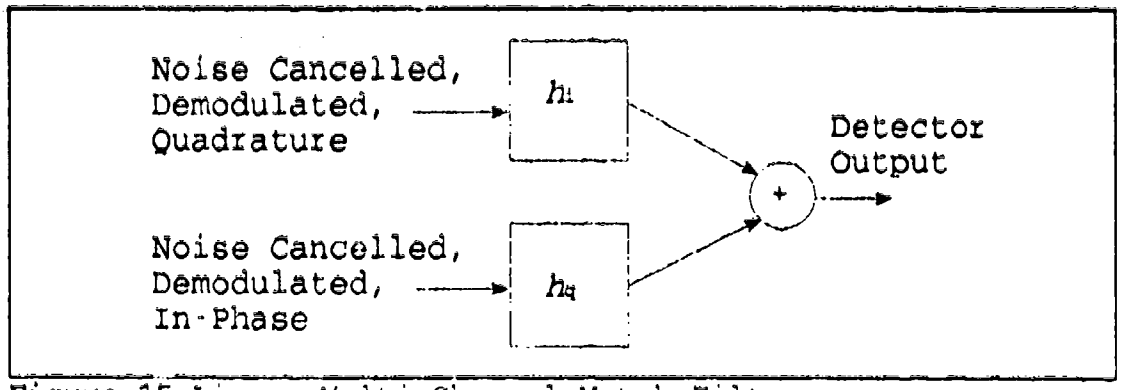


Figure 15 Linear Multi-Channel Match Filter

located in the Appendix. The detector was not used by the author in this work relating to the analysis of the SER algorithm, but it is presented as complementary work relating to the overall project.

E. SIGNAL LOCATION

The signal strength analysis showed that the signal divided by 2000 is easily discernable from the background at the output of the detector. The effect of varying the location of the signal will allow analysis to get a better feel for the expected SNR based on a larger sample size for averaging than previously used. A second benefit of this analysis is to see the effect of the initialization process on the SNR. The signal will be shifted by a quarter of its length for each run, or 2048 points. The SNR is plotted versus the location of the center of the signal in Figure 16.

The data is run with and without the SER algorithm to get a good feel for the enhancement achieved with the SER

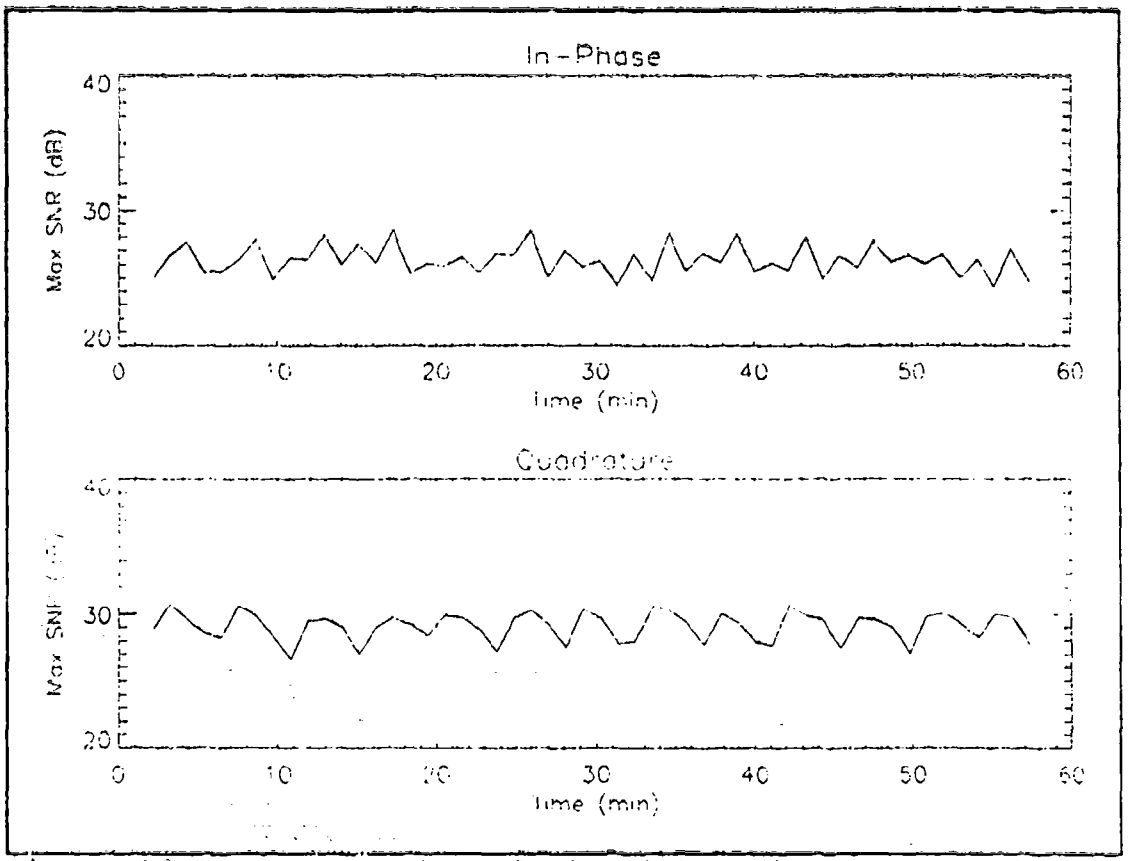


Figure 16 SNR VS Location of Signal In Noise Vector (Y)

algorithm. Three separate locations of the signal will be used for each run to reduce calculation time. The use of multiple signals causes no effect on the SNR calculation as long as the signals are not allowed to overlap.

VII. RESULTS AND CONCLUSION

A. ANALYSIS OF RESULTS

The use of the SER algorithm enhances the ability to detect narrowband low power signals buried in the background noise. Figure 17 shows the spectrum of the noise sample from sensor XE before and after the use of the SER algorithm. The frequency range is from one to ten hertz, which is in the range of concern for the JHU project. The spike seen at four hertz (note the x-axis is a logarithmic scale) is caused by the processing equipment. This is a man-made noise source and will be filtered out in the final project design. The figure illustrates the 10-15 dB of noise cancellation possible by using this method of adaptive noise cancellation.

The location of the signal in comparison to the start of the adaptive process has been shown to have very little effect on the signal enhancement. The system implementation will most likely cause the algorithm to only require initialization once a week and operate constantly otherwise. The initialization process takes about 100 iterations or 1.67 minutes before settling out but is heavily data dependent. Even while settling, the performance of the algorithm is adequate to receive most signals. For the purpose of this project, initialization should be of no concern.

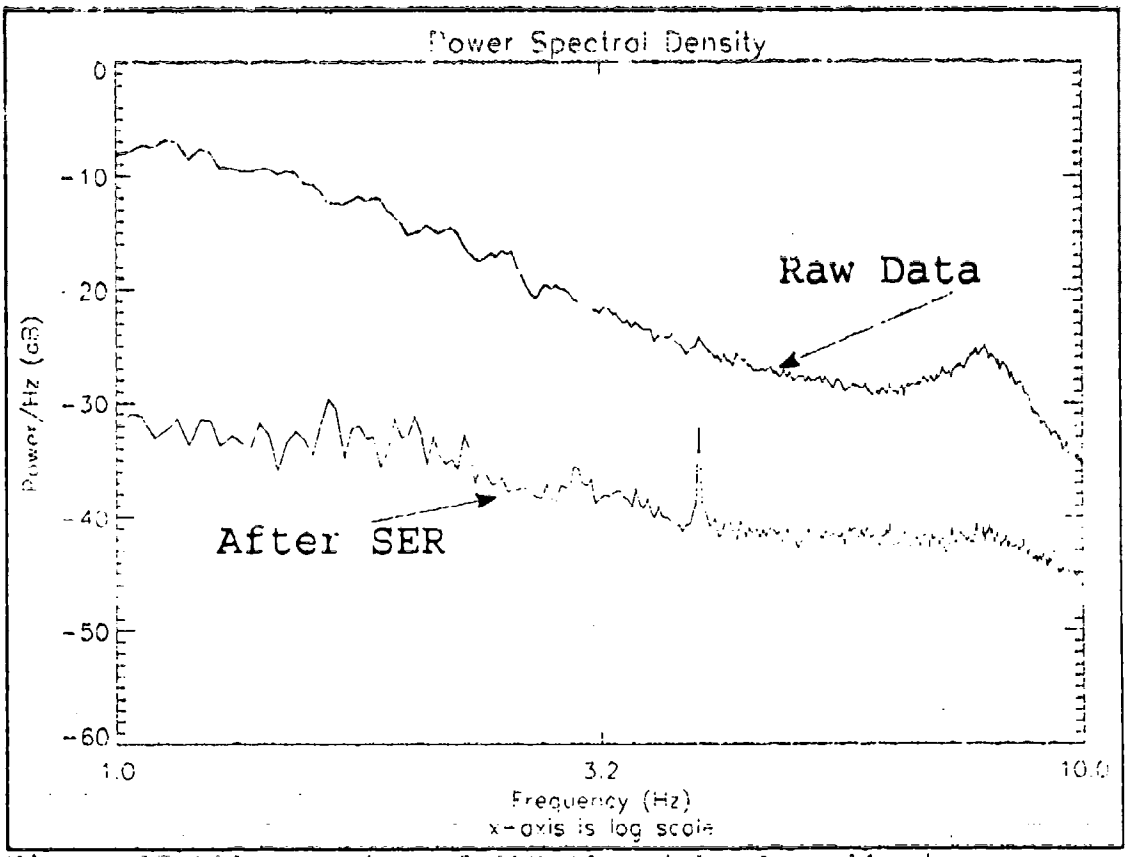


Figure 17 Illustration of SER Algorithm Cancellation Capability, Power Spectral Density of 1-10 Hz

B. CONCLUSIONS AND RECOMMENDATIONS

It was hoped at the beginning of the research that information could be gained on the characteristics of the noise environment. The stationarity of the background noise environment was found to be at least nine seconds. This is believed to be only a characteristic of the specific data set that is used. The length of stationarity affects the value of the forgetting factor, α, and is implied by the use of the SER algorithm. A better feel for a good average value to use for the length of stationarity will result from the use of more

data sets. This analysis needs to be conducted and could be a source of further work in this area.

A second area of work that could develop from this research is a study of the cancellation possible from a vertical above-surface sensor and a submerged horizontal sensor. The purpose of this would be to differentiate between surface and submerged activity. This may prove to be impossible but it would be interesting to see the amount of correlation possible between the two locations. This type of arrangement could also lead to a wider distance between sensors which could also increase the knowledge of the background environment.

The purpose of the research presented in this thesis is to develop a method to adaptively cancel the ELF noise present in a submerged sensor array. The SER algorithm appears to be an adequate method to accomplish this goal of noise cancellation. The cancellation ratio in the frequency range of interest is, on average, 12.5 dB. This cancellation ratio along with the use of optimal detectors will assist the system in detecting narrowband signals buried deep in the background noise.

APPENDIX

The second detector used in conjunction with this project of measuring the geoelectric background noise is an optimal nonlinear envelope detector. The symbolism used is slightly different but efforts are made to define all relevant terms. The in-phase and quadrature data (after noise cancellation) are denoted by r_i^i and r_i^q, respectively. Similarly the in-phase and quadrature noises are represented by n_i and n_i. Assuming a joint Gaussian distribution we have the following two hypothesis:

$$H_0 : (\boldsymbol{r}^i , \boldsymbol{r}^q) = (\boldsymbol{n}^i , \boldsymbol{n}^q)$$

$$H_1 : (\boldsymbol{r}^i , \boldsymbol{r}^q) = (\boldsymbol{n}^i , \boldsymbol{n}^q) + (\boldsymbol{s}^i , \boldsymbol{s}^q)$$

where H_0 denotes the null hypothesis (noise only) and H_1 denotes the alternative (signal and noise present). The vector notation means

$$\boldsymbol{r} = [\boldsymbol{r}^i, \boldsymbol{r}^q] = [(\boldsymbol{r}_1^i, \boldsymbol{r}_1^q) , \ldots , (\boldsymbol{r}_N^i, \boldsymbol{r}_N^q)]$$

where N denotes the length of the interval for which the decision is being made. It is then clear that the probability density function for the received data is as follows:

$$p(\boldsymbol{r}|H_1) \propto \exp\left[-\frac{1}{2} (\boldsymbol{r}-\boldsymbol{s})^T \underline{R}_{nn}^{-1} (\boldsymbol{r}-\boldsymbol{s})\right]$$

where $\underline{R}_{nn}$ denotes the covariance matrix of the noise vector. In this particular case, it is assumed that the in-phase and

the quadrature components have been decorrelated and whitened with the same variance σ^2 so

$$p(\boldsymbol{r}|H_1) = \left(\frac{1}{2\pi\sigma^2}\right)^N \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{k=1}^{N}(r_k^i - s_k^i)^2 + \sum_{k=1}^{N}(r_k^q - s_k^q)^2\right]\right\}$$

Introducing polar variables (r_k^i , ϕ_k^i) and (r_k^q , ϕ_k^q) it is found that

$$p(\boldsymbol{r}|H_1) = \frac{\prod_{k=1}^{N} r_k}{2\pi\sigma^{2N}} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{k=1}^{N} r_k^2 + \sum_{k=1}^{N} s_k^2 - 2\sum_{k=1}^{N} r_k s_k \cos(\phi_k)\right]\right\}$$

where $r_k^2 = r_k^{i2} + r_k^{q2}$ and $\tan(\phi_k) = r_k^q / r_k^i$. Integrating the above over the angular variables will give

$$p(\boldsymbol{r}|H_1) = \frac{\prod_{k=1}^{N} r_k}{2\pi\sigma^{2N}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{N}(r_k^2 + s_k^2)\right\}\prod_{k=1}^{N} I_0\left(\frac{s_k r_k}{\sigma^2}\right)$$

where I_0 is the modified bessel function of order zero. The vector $\boldsymbol{r}$ on the left hand side now only refers to the various envelopes, $\boldsymbol{r} = [\, r_1, \ldots, r_N]$. The same is true for the signal envelope. The likelihood ratio detector is found to be:

$$\lambda = \frac{p(\boldsymbol{r}|H_1)}{p(\boldsymbol{r}|H_0)} = \exp\left[-\frac{1}{2\sigma^2}\sum_{k=1}^{N} s_k)\right]\prod_{k=1}^{N} I_0\left(\frac{s_k r_k}{\sigma^2}\right)$$

The log-likelihood ratio is then:

$$\log\lambda = -\frac{1}{2\sigma^2}\sum_{k=1}^{N} s_k + \sum_{k=1}^{N}\log\left\{I_0\left(\frac{s_k r_k}{\sigma^2}\right)\right\}$$

The first term is a constant, dependent only on the known signal, and can be dropped. Finally the test statistic or detector output is[Ref 10]:

$$\log\lambda = \sum_{k=1}^{N} \log\left\{I_0\left(\frac{s_k r_k}{\sigma^2}\right)\right\}.$$

The implementation and analysis of this detector scheme is left as a possible follow-on thesis topic.

LIST OF REFERENCES

1.Lokken,J.E., "Instrumentation For Receiving Electromagnetic Noise Below 3,000 CPS," paper presented in Proceeding of a NATO Advanced Study Institute, Bad Homburg, Germany, 22 July-2 August 1963.

2.Galejs,J., "Terrestrial Extremely Low-Frequency Propagation," paper presented at the NATO Advanced Study Institute, Bad Homburg, Germany, 22 July - 2 August 1963.

3.Brock-Nannestad,L., "Electromagnetic Background Noise in the ELF Range," paper presented at the Symposium of the Avionics Panel of AGARD, 12th, Paris, France, 25-29 April 1966.

4.Cheng, D. K., *Field and Wave Electromagnetics*, p. 586, Addison-Wesley Publishing Company, 1990.

5.Nichols, E.A. and Morrison, H.F., "Signals and Noise in Measurements of Low-Frequency Geomagnetic Fields," *Journal Of Geophysical Research*, v. 93, pp. 13743-13754, 10 November 1988.

6.Haykin, Simon, "Introduction to Adaptive Filters", pp.1-23, Macmillan Publishing Company, New York, 1984.

7.Widrow, B. and Stearns, S.D., *Adaptive Signal Processing*, Prentice-Hall, Inc., 1985.

8.Lee, R. C., *Optimal Estimation, Identification, and Control*. Cambridge, Mass.:MIT Press, 1966, Sec 4.3.1.

9. Private Communication, A. Najmi, Senior Professional Staff, STR Group, Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland, 10 August 1993.

10. Whalen, A.D., *Detection of Signals in Noise*, pp.170-175, Academic Press, Inc., 1971.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center — 2
   Cameron Station
   Alexandria, Virginia 22304-6145

2. Library, Code 52 — 2
   Naval Postgraduate School
   Monterey, California 93943-5002

3. Director, Space And Electronic Combat Division (N64) — 1
   Chief of Naval Operations
   Washington, District of Columbia 20350-2000

4. Chairman, Electronic Warfare Academic Group — 1
   Code EW
   Naval Postgraduate School
   Monterey, California 93943-5002

5. McDonough, Robert N. (8-368A) — 1
   Submarine Technology Department
   Johns Hopkins University Applied Physics Laboratory
   Johns Hopkins Road
   Laurel, Maryland 20723-6099

6. Bjerkaas, Allan W. (25-266) — 1
   Submarine Technology Department STR Group
   Johns Hopkins University Applied Physics Laboratory
   Johns Hopkins Road
   Laurel, Maryland 20723-6099

6. Hart, Lynn W. (25-260) — 1
   Submarine Technology Department STR Group
   Johns Hopkins University Applied Physics Laboratory
   Johns Hopkins Road
   Laurel, Maryland 20723-6099

7. Najmi, Amir-Homayoon (25-246) — 1
   Submarine Technology Department STR Group
   Johns Hopkins University Applied Physics Laboratory
   Johns Hopkins Road
   Laurel, Maryland 20723-6099

8. Farley, Danny G. (Code 3A) — 1
   Electronic Warfare Academic Group
   Naval Postgraduate School
   Monterey, California 93943-5002